Agglomeration of needle-like Crystals in Suspension – I. Measurements †

David R. Ochsenbein,[‡,‖] Thomas Vetter,[¶,‖] Stefan Schorsch,[§] Manfred Morari,[‡] and Marco Mazzotti[*,§]

[1] *Automatic Control Laboratory, ETH Zurich, Physikstrasse 3, CH-8092 Zürich, Switzerland, University of Manchester, School of Chemical Engineering and Analytical Science, M13 9PL Manchester, United Kingdom, and Institute of Process Engineering, ETH Zurich, Sonneggstrasse 3, CH-8092 Zürich, Switzerland*

E-mail: marco.mazzotti@ipe.mavt.ethz.ch

Phone: +41 44 632 24 56. Fax: +41 44 632 11 41

[2] **Abstract**

[3] A technique for the detection and measurement of the agglomeration of needle-like
[4] particles is presented. A novel image analysis routine, based on a supervised machine
[5] learning strategy, is used to identify agglomerates that are subsequently characterized
[6] by their volume. Through repeated measurement of a large number of agglomerates, a

†NOTICE: this is the author's version of a work that was accepted for publication in Crystal Growth & Design. Changes resulting from the publishing process, such as editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Crystal Growth & Design , doi: 10.1021/acs.cgd.5b00094

*To whom correspondence should be addressed

‡Automatic Control Laboratory, ETH Zurich, Physikstrasse 3, CH-8092 Zürich, Switzerland

¶University of Manchester, School of Chemical Engineering and Analytical Science, M13 9PL Manchester, United Kingdom

§Institute of Process Engineering, ETH Zurich, Sonneggstrasse 3, CH-8092 Zürich, Switzerland

‖These authors contributed equally to this work.

1D particle size distribution of agglomerates is reconstructed. Concurrently, established tools are used to characterize needle-like primary crystals, whose shape is described by cylinders and whose population can be described by a separate, two-dimensional particle size and shape distribution. The performance of the classifier is evaluated and the reproducibility of the measurement demonstrated for the case of β L-glutamic acid. For the same system, the agglomeration behavior is studied for varying operating conditions and general trends are analyzed.

1 Introduction

The study of agglomeration in particulate systems has received considerable attention for a number of important applications. Many publications deal with nano-sized particles, where agglomeration is a main contributor to size enlargement and therefore has a considerable impact on particle properties (see, e.g. Limbach et al.[1], Yuk et al.[2]). In the case of crystallization from solution, agglomeration has been investigated experimentally in particular for the precipitation not only of (slightly soluble) salts,[3–8] but also for organic compounds.[9–13] In these cases, agglomeration is often studied due to its negative effects on product properties, such as the broadening of the particle size distribution (PSD) and the entrapment of solvent.[14–16] Therefore, strategies to avoid agglomeration are often sought.[9] However, occasionally agglomeration may also be exploited, e.g., when spherical particles that ease downstream processing are produced.[10,16]

Hitherto, the detection and quantification of agglomerates and of the degree of agglomeration have been accomplished both indirectly and directly, i.e., through measurements of the total particle size distribution or through imaging methods, respectively. In the case of the various microscopy methods, research has often been based on human assessments and it is only in more recent times that more quantitative and automated techniques for description and detection of agglomerates have been actively pursued.[17–20] In this regard, agglomeration studies share a common evolution with the study of the morphology of primary crystals. The

latter has seen considerable progress in the last few years, with the advent of an era of new imaging methods, capable of measuring both size and shape of particles in real processes.[21–24] These tools allow for an unprecedented level of detail regarding the interpretation of images and characterization of systems[21,25,26] and it seems sensible to assume that this progress may bring forth new discoveries in the field of agglomeration.

Given the highlighted similarities and the mutual importance of the phenomena for crystallization processes, it is conspicuous that (primary particle) shape and agglomeration are very seldom studied together. Indeed, while the habit of agglomerates under varying conditions has been the subject of studies in the past, the habit of the primary particles constituting these agglomerates has received much less attention.[11,27–30] With the notable exception of the work by Briesen[31], to the knowledge of the authors, no work studying primary particle shape in agglomerates or the effect of morphology on agglomeration exists in the literature for crystals in suspension. Yet, such an analysis is of great practical interest, as it may further support (or falsify) the notion that knowledge and control of crystal morphology is crucial to establish a complete description and ultimately an effective control of crystallization.

In this work, a novel way of identifying and characterizing agglomerates from images is introduced. Together with previously created tools to measure the size and shape of non-equant shaped particles, this allows for the simultaneous reconstruction of both the size distribution of agglomerates as well as the primary particle size and shape distribution. The reproducibility of the obtained measurements is investigated for the case of β L-glutamic acid and the behavior of this system is studied for varying operating conditions.

This work is structured as follows: Section 2 outlines the experimental setup used in this work. Section 3 introduces the novel image analysis procedure and data interpretation strategy, i.e., the generation of a training set and the calculation of particle descriptors that are then used to automatically classify particles as agglomerates or primary particles based on a nonlinear support vector machine. In Section 4, the reproducibility of the proposed measurements is studied for the case of β L-glutamic acid under varying process conditions before the

qualitative behavior of the system is discussed in detail.

2 Experimental setup and procedures

2.1 Setup

All experiments were carried out in a temperature-controlled 500 mL jacketed glass crystallizer that was previously described in detail.[12] In order to characterize the evolution of the crystallization process, the crystallizer was equipped with an *in situ* ATR-FTIR probe, an *in situ* FBRM probe and was connected to an *ex situ* stereoscopic imaging device through a sampling loop, described in detail in the following.

2.2 Characterization techniques

2.2.1 ATR-FTIR and FBRM Measurements

To monitor the liquid phase concentration of L-glutamic acid, an ATR-FTIR spectrometer (ReactIR 45m with a fixed arm mirror conduit, Mettler Toledo Switzerland) was used. Recorded spectra were preprocessed by mean centering and considering the first derivative. A calibration model at process temperature and a supersaturation range between $S_0 = 1$ and $S_0 = 5$ covering the entire experimental range at multiple wavelengths was built using partial least squares as detailed elsewhere.[25,32] Note that the supersaturation is defined here as

$$S(t) = \frac{c(t)}{c^*(T)} \tag{1}$$

where c is the solute concentration, $c^*(T)$ is the temperature dependent solubility of β L-glutamic acid and $S_0 = S(t = 0)$.[33]

Furthermore, during some of the experiments, the chord length distribution (CLD) was monitored using an *in situ* FBRM probe (Lasentec 600L, Redmond, USA) in order to detect the occurrence of nucleation in the system. However, due to repeated fouling of the sensor at

high supersaturations (i.e., precisely in the cases where nucleation is most likely), the CLD measurements could not be utilized to conclusively confirm or disprove the occurrence of nucleation.

2.2.2 Imaging Device

Online imaging of particles in suspension was performed using a measurement device in which particles are photographed from perpendicular directions in a transparent flow through cell with optimal optical properties. The suspension is fed to the cell and recycled into the reactor using a sampling loop and the resulting stereoscopic images are treated with global thresholding. The interested reader can find all the details of the experimental setup in our previous publications.[23,24] Subsequently, connected component extraction delivers the contour of all objects for both camera views and a matching algorithm assigns each object in the first photograph to the corresponding projection in the second image through a comparison of the position of the boundary points along the common coordinate. This procedure exhibits an important advantage over monoscopic imaging solutions, not only as it overcomes the orientation dependence, but because it also allows to distinguish between particles that are indeed in contact or agglomerated and those that are simply overlapping in only one of the two images.

2.3 Experimental procedures

2.3.1 Preparation of seed crystals

Seed crystals of the β polymorph of L-glutamic acid have been prepared in a two step process by first precipitating the α polymorph through a pH-shift reaction of equimolar amounts of hydrochloric acid (Fluka, 37%) and glutamic acid monosodium salt hydrate (Sigma Aldrich, purity $> 99\%$) in deionized water (obtained from a MilliQ Advantage A10 system (Merck Chemicals)) at 5°C and stirring continuously for one hour. Particles were filtered off and added to a saturated (with respect to the α form) aqueous solution at 45°C

in agreement with previous studies.[32] After the transformation of the prismatic α-crystals to needle-like β-crystals had occurred, the β polymorph product was filtered off and wet-sieved. Characteristics of the obtained seed distributions are given in Table 3.

2.3.2 Agglomeration experiments

The experimental solution was prepared by mixing hydrochloric acid and a clear solution of the monosodium salt in water to obtain the desired initial supersaturation. The reactor was kept at constant temperature and at a desired stirring rate before seed crystals of the β polymorph were added in order to start the agglomeration experiments, which were typically conducted until the solute concentration had reached quasi-equilibrium (see Figure 1). An overview of the experimental conditions is reported in Table 2. It should be noted that the suspension density of all experiments was chosen such that the volume fraction of particles remains moderate throughout the experiments, with values ranging from 0.4% at the beginning to maximally 5% in the end of some experiments. Broadly speaking, the effect of particle-liquid interactions on the behavior of the fluid becomes increasingly important for increasing values of the volume fraction and care should be taken when trying to extrapolate findings or parameters.[34,35]

Note also that at the high supersaturations used in these experiments, the metastable α polymorph of L-glutamic acid is supersaturated as well. However, due to the presence of considerable amounts of β seed crystals and fast desupersaturation due to crystal growth and agglomeration, the formation of α crystals is highly unlikely. In addition, the absence of needle breakage was verified by monitoring the average length of particles in a saturated solution using the highest stirring rate used in the agglomeration experiments. As shown in Figure 2, the average length is constant even under prolonged stirring.

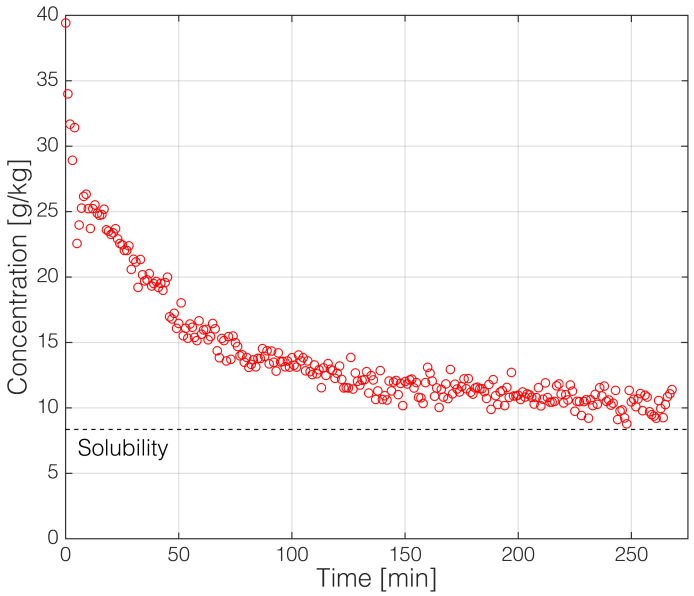

Figure 1: Concentration vs. time for Experiment 1 (cf. Table 2). The dashed line indicates the solubility at 25°C.

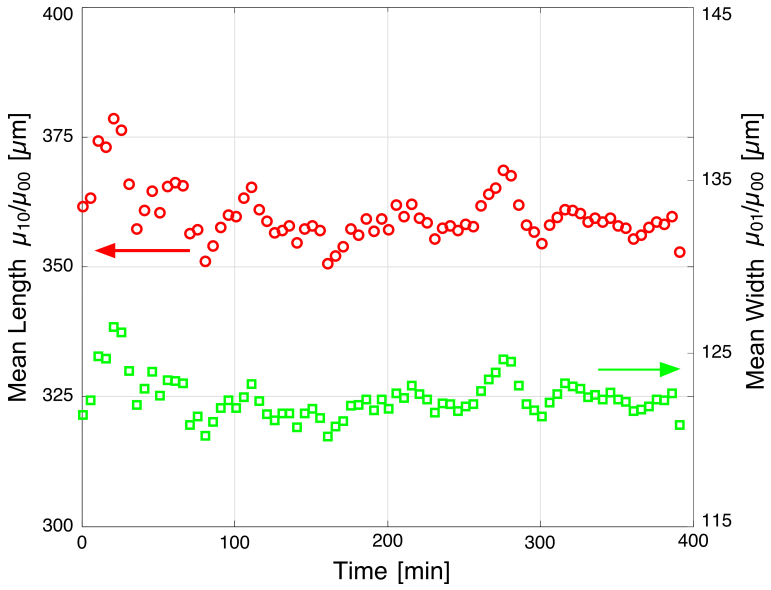

Figure 2: Measurement of seed particles in a saturated solution ($S = 1$) over 6.5 hours. A decrease in the number-weighted, average length and width of the needles, indicating breakage, was not observed.

3 Image Analysis

The key step for the successful analysis of agglomeration is to obtain knowledge of the particle size distribution(s) and of the degree of agglomeration. In order to obtain this data, the stereoscopic imaging device described previously was used. It has already been demonstrated that this setup allows for the accurate monitoring of size and shape of ensembles of crystals. However, the use of such a setup to distinguish between primary particles and agglomerates and to estimate the size of such agglomerates is novel. Combining these efforts shall make the measurement of quantitative, multidimensional particle size distributions for the primary particles and one-dimensional particle size distributions for the agglomerates possible. A schematic of the entire PSD reconstruction procedure, including the imaging and classification steps, is presented in Figure 3. In the following, the generation of the training set, which is used both for a benchmark analysis as well as the foundation for a supervised learning strategy, is reported first, followed by an outline of the new methods. Throughout the text, the existing image analysis scheme will only be briefly summarized, while its modifications, which have been introduced in the course of this work, will be described in greater detail.

3.1 Training Set Generation

Clearly, the drastic shape change between primary particles (needles) and agglomerates in the case of β L-glutamic acid makes image analysis a promising method for identifying agglomerates. However, on the one hand, the efficacy of automated image analysis depends on the type and the quality of the measurements, while on the other hand, an unequivocal identification of agglomerates from a small number of images is, occasionally, simply not possible. An assessment of the maximum achievable identification performance prior to creating an image analysis routine is therefore in order. We have decided to do this by comparing evaluation results for a given set of images obtained from different human experts. The agreement can be used as a measure of the expected identification performance of

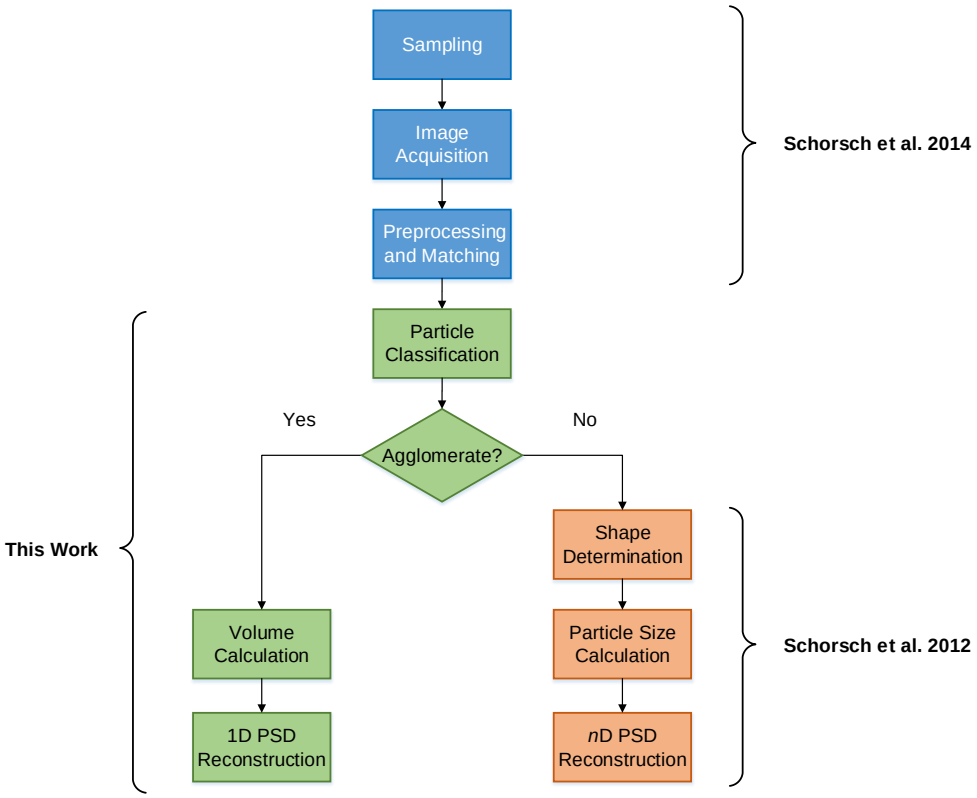

Figure 3: Schematic overview of sampling and imaging analysis procedure including references indicating detailed sources for the individual steps.

an automated system based on image analysis. Furthermore, the set of classified particles subjected to the evaluation of human experts can then be used as training set for an automated, supervised learning strategy.

In order to obtain such an identification performance benchmark, a graphical user interface (GUI) was developed that displays photographed particles and that allows human users to assign the observed objects to different groups, i.e., either primary particles or agglomerates. A screenshot of this GUI is presented in Figure 4. Four crystallization experts were asked to classify 500 seed and product particles, based on the two orthogonal projections obtained through the flow through cell (FTC) measurement device. Seed particles were grown under conditions where a low number of agglomerates is expected, while the product crystals were grown under conditions which favor agglomeration. The experts proceeded to classify the

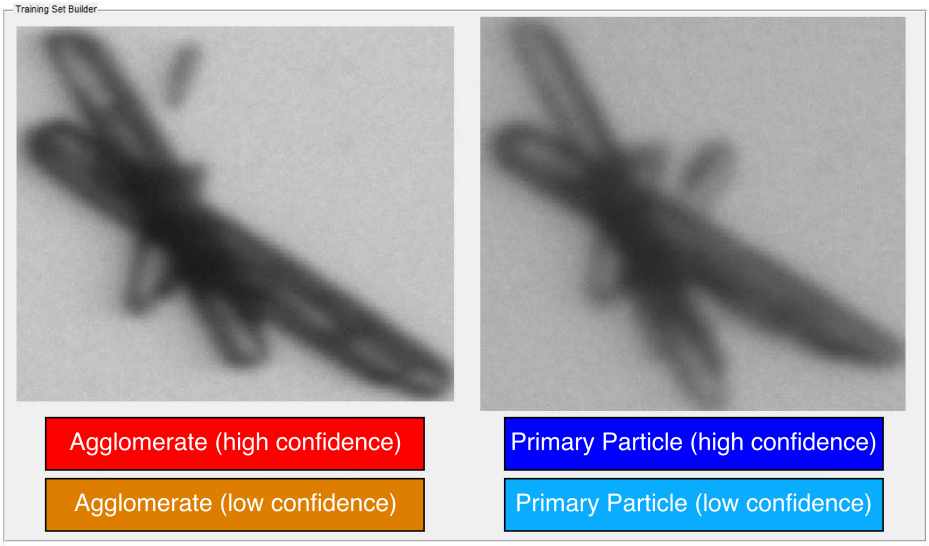

Figure 4: Screenshot of the graphical user interface used for benchmarking and training set creation. Users are provided with the two projections of particles obtained using the stereoscopic imaging setup and asked to choose the particle class and confidence level.

particles into four classes: (P1) primary particle (high level of confidence), (P2) primary particle (low confidence), (A1) agglomerate (high confidence) and (A2) agglomerate (low confidence).

In order to gauge whether the number of analyzed particles is sufficient, the computed volume weighted degree of agglomeration, defined as the ratio of total agglomerate volume over total volume of agglomerates and primary particle (cf. Section 4 for the formal definition), as a function of the number of evaluated particles is plotted for all experts in Figure 5. As shown in the plot, the absolute difference between the evaluated degree of agglomeration after 300 samples and the final value (after 500 samples) for each individual expert is typically less than 5%, thus indicating that the set size is sufficient for our purposes. However, Figure 5 also reveals significant discrepancies in the evaluation results of the different experts. A more detailed analysis is possible based on the classification results reported in Figure 6, which considers the seed and product crystals separately, as well as all crystals combined.

Two observations can be made immediately: the amount of particles classified as agglomerates (classes A1 and A2) increases from seeds to products; the (self-assessed) confidence in the

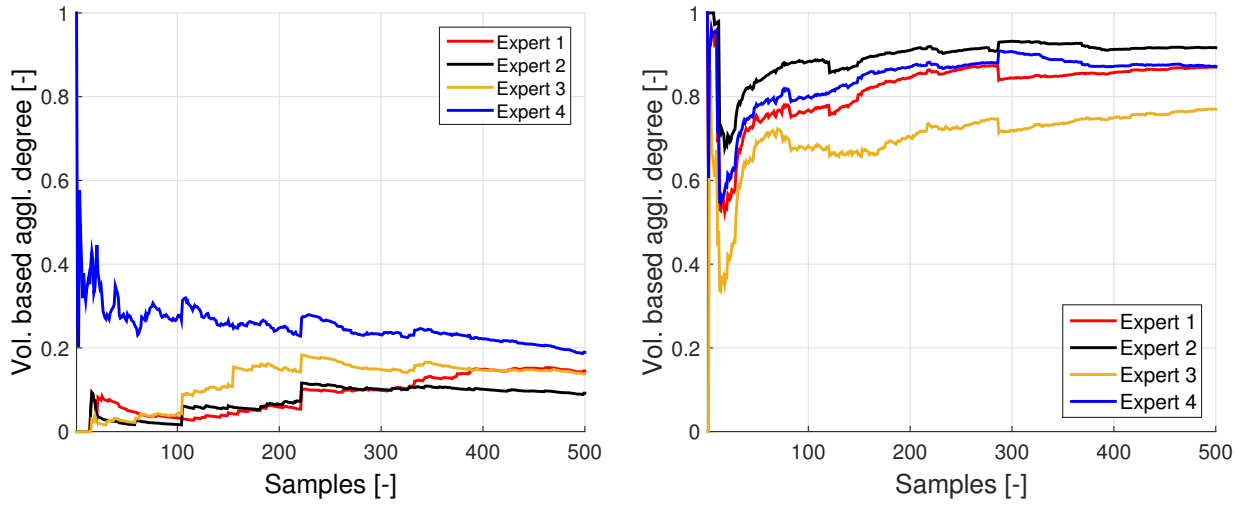

Figure 5: Evolution of the agglomeration degree over the sample number when analyzed by experts for (a) seeds and (b) products.

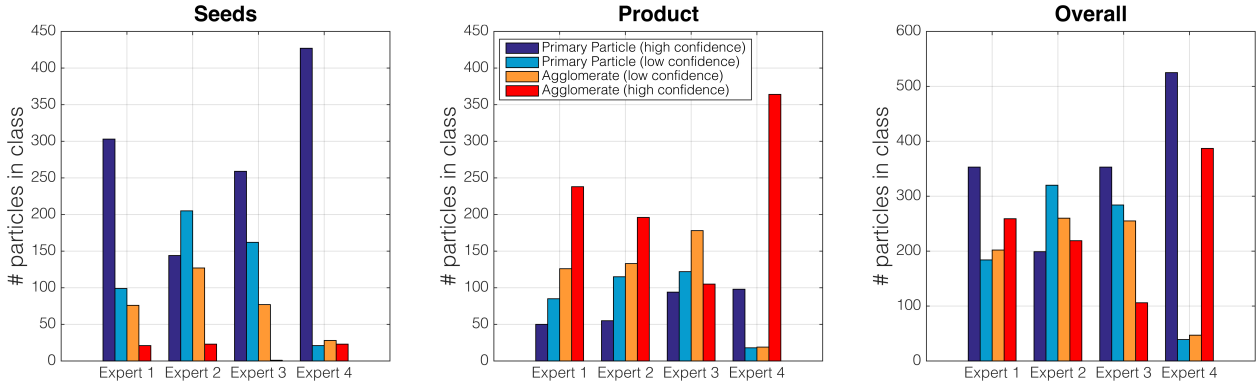

Figure 6: Summary and comparison of expert evaluation for seed, product and total of investigated particles.

classification carried out by each expert varies widely. Indeed, the ratio of highly confident to less confident decisions varies from approximately 40:60 to approximately 90:10 between the experts. This indicates that the interpretation of images obtained with the given setup carries a degree of ambiguity that might also affect an automated classification algorithm. In order to mitigate this issue, for the remainder of the analysis, it was chosen to merge classes (P1) and (P2) into a single primary particle class (with label -1), and classes (A1) and (A2) into an agglomerate class with label (1), i.e., the confidence levels with which the decisions were taken are neglected.

Table 1: Agreement between experts and the trained computer in distinguishing between agglomerates and primary particles.

	Computer	Expert 1	Expert 2	Expert 3	Expert 4
Computer	100%	-	-	-	-
Expert 1	82%	100%	-	-	-
Expert 2	73%	77%	100%	-	-
Expert 3	79%	81%	75%	100%	-
Expert 4	86%	84%	76%	81%	100%

After applying this simplification, the agreement between each pair of experts can be quantified in the form of the binary agreement measure, i.e., by calculating the fraction of cases where the two experts have chosen the same class; the corresponding values are reported in Table 1. As shown in the table, the agreement between experts is reasonable, though not perfect, as was to be expected. A more thorough investigation of the statistical significance of the results is provided in the supplementary material. Nevertheless, the ultimate purpose of the expert data is to be used as robust training set for the machine learning classifier. To achieve such a robust starting point, only the subsets of particles unequivocally identified by *all* experts as either class are considered as the training set for the computer. Adding this constraint reduces the cardinality of the training set to 609 particles (out of the initial 1,000 objects; consisting of 249 agglomerates and 360 primary particles), which is still a sufficiently large number to develop a meaningful classifier.

3.2 Particle Classification

The findings of Section 3.1 were taken as a benchmark for the development of an automated image analysis routine which bases its classification on a small set of measurable quantities. Numerous possible descriptors for agglomerate detection and description have been used in the literature, and a proper choice is clearly decisive.[18,30,31,36] In this work, the selection of features was guided by two considerations. First, all descriptors should be scale-independent and hence dimensionless. Second, the number of extracted quantities should be small, so

as not to overfit the data set. After careful examination of the different options, three dimensionless, characteristic measures (per projection) were chosen as features to be used in the classification step. These are (1) the circularity, defined as the projection area over area of circle with the same average distance from centroid to boundary

$$I_j^{\mathrm{CIR}} = \frac{A_{\mathrm{proj},j}}{\pi r_{j,\mathrm{mean}}^2} \tag{2}$$

where $r_{j,\mathrm{mean}}$ is the average distance between contour and area center of a projection and $A_{\cdot}$ denotes the corresponding area (the index $j \in \{1, 2\}$ identifies the camera source of the feature); (2) the convexity (sometimes referred to as solidity), defined as the pixel area divided by the area of the convex hull

$$I_j^{\mathrm{VEX}} = \frac{A_{\mathrm{proj},j}}{A_{\mathrm{conv},j}} \tag{3}$$

and (3) the concavity index, here defined as ratio between the area of the largest concavity part of the projection and the area of the projection

$$I_j^{\mathrm{CAV}} = \frac{A_{\mathrm{maxCAV},j}}{A_{\mathrm{proj},j}} \tag{4}$$

Illustrations for the latter two indices are provided in Figure 7. Given that two projections from orthogonal directions are obtained, each particle can be associated to a six-dimensional feature vector x, such that

$$\mathsf{x} = \left[I_1^{\mathrm{CIR}},\, I_2^{\mathrm{CIR}},\, I_1^{\mathrm{VEX}},\, I_2^{\mathrm{VEX}},\, I_1^{\mathrm{CAV}},\, I_2^{\mathrm{CAV}}\right]^{\mathsf{T}} \tag{5}$$

The feature vectors of all $N = 609$ are then computed and used for the design of an automated classification algorithm that is based on a supervised learning scheme. In other words, based on the training data set $\mathcal{D}$

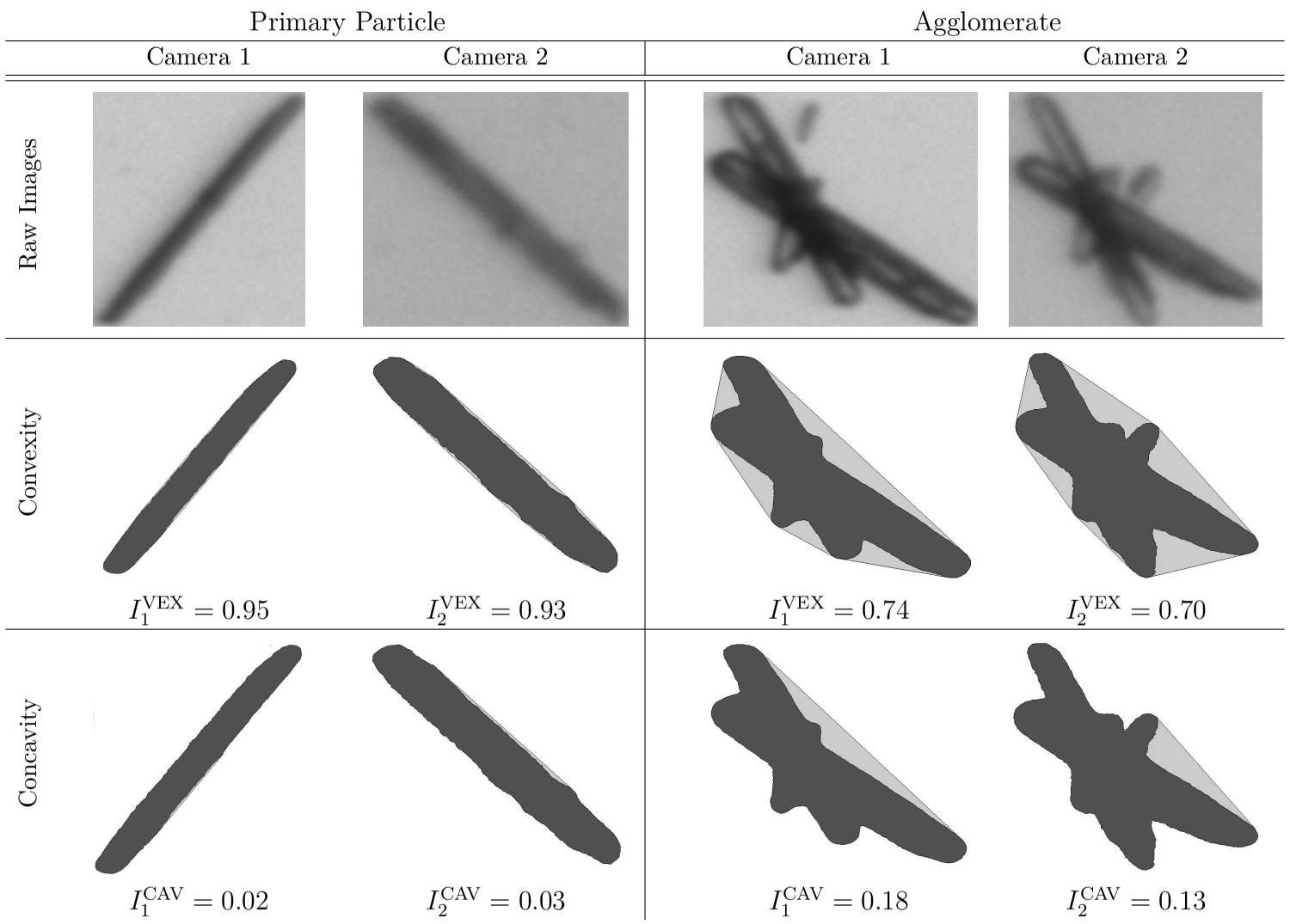

Figure 7: Two particles as observed by the orthogonal camera setup. The left particle would be classified as primary particle, the right particle as agglomerate. The top row contains raw images. The second row shows the convexity, i.e., the dark area is the projection of the particle, while the bright area is the convex hull. Convexity I_j^{VEX} is defined as ratio between particle area over area enclosed by the convex hull. The bottom row shows the concavity, i.e., the dark area is again the particle projecton, while the largest concavity is shown as bright area. The ratio of the two is defined as I_j^{CAV}.

$$\mathcal{D} = \left\{ \left(\mathsf{x}^{(i)}, y^{(i)} \right) \mid \mathsf{x}^{(i)} \in \mathbb{R}^6,\ y^{(i)} \in \{\text{-}1, 1\},\, i = \{1, \ldots, N\} \right\} \tag{6}$$

for which the labels are known, we aim at finding some rule or logic that identifies agglomerates with high accuracy. This rule can then be used for the automatic classification of a new particle, i.e., to predict the label $\hat{y}$ given a new feature vector x.

Before discussing the classifier, it is instructive to project $\mathcal{D}$ onto planes spanned by the descriptors, as this allows to get some intuition regarding the contribution of individual features to separate the data; the fifteen possible combinations are visualized in Figure 8. It is clear that in all cases the classes of agglomerates (red circles) and primary particles (black crosses) are partially overlapping and no straightforward boundary between the two groups can be drawn. This indicates that a simple combination of two descriptors is insufficient to distinguish between agglomerates and primary particles. However, as is evident from the trace plots of I_1^{VEX} vs. I_2^{VEX} (top row, first plot from the left), I_1^{CIR} vs I_2^{CIR} (middle row, last plot) and I_1^{CAV} vs I_2^{CAV} (bottom row, last plot), all descriptors provide at least *some* separation between the two particle classes.

As described in the supplementary material accompanying this text, several methods for the classification have been tested, including linear and quadratic discriminant analysis and a linear support vector machine (SVM). Alas, the separation performance for these classifiers was found to be insufficient. However, focusing on the SVM method, which aims at finding the maximum-margin hyperplane separating the data, a common solution to enhance separability is to apply the so-called *kernel trick*, i.e., to perform a nonlinear mapping of the training data followed by an attempt to (linearly) separate classes in the transformed feature space.[37] The benefit of the method is that the classifier in the new space may correspond to a *nonlinear* separation in the original space.

The soft-margin implementation of such a nonlinear support vector machine, which allows for a certain degree of inseparability, was used in this work, employing a radial basis function as kernel. The method, which contains two tuning parameters, has been implemented via

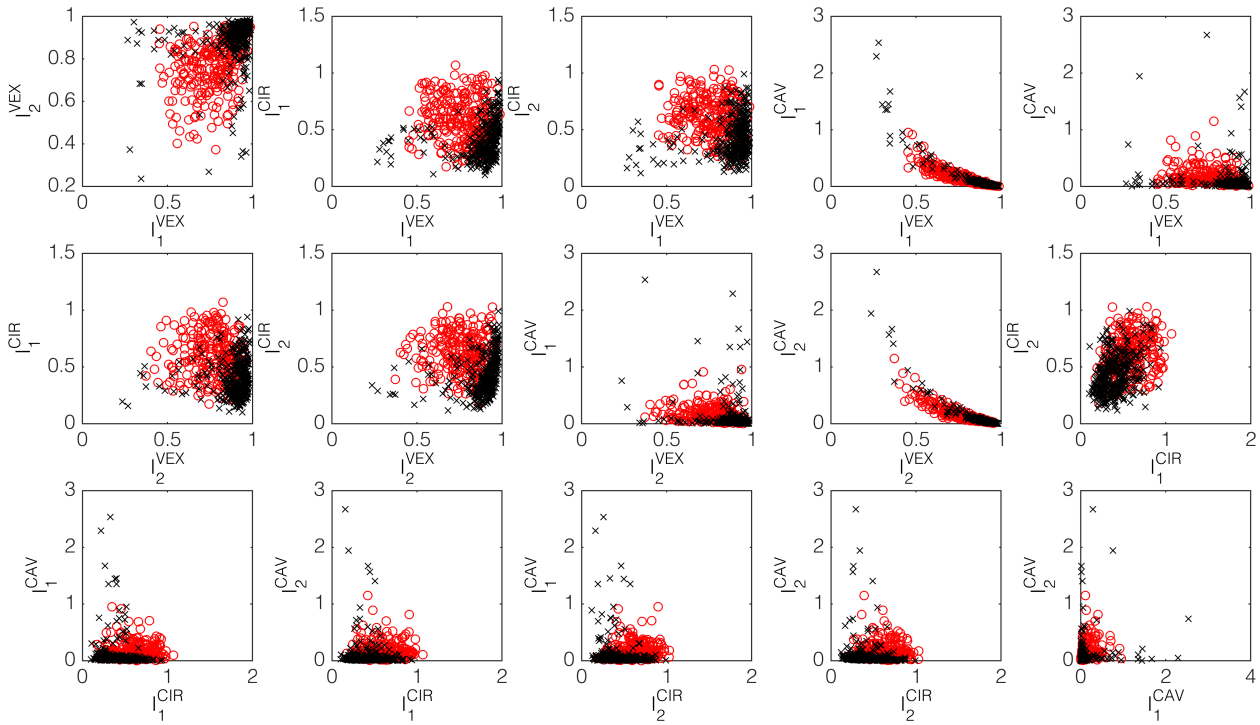

Figure 8: Trace plots for combinations of two descriptors. The red circles and black crosses represent particles unanimously classified by the experts as agglomerates and as primary particles, respectively.

Matlab (using the machine learning toolbox) and the final training and prediction errors have been calculated as 7.9% and 8.2%, respectively. The training error refers to the percentage of particles in the training set that are misclassified by classifier, while the prediction error has been obtained by using a standard 10-fold cross validation procedure. In such a procedure the training data is randomly partitioned into 10 equal-sized segments, 9 of which are used to train the classifier, while the remaining one is used to assess the quality of the prediction. This procedure is then repeated 10 times and the mean value of the prediction error is reported. The agreement between the machine classifier and the individual experts, reported in Table 1, was also found to be acceptable, i.e., comparable to the other values. While many other kernel functions could be used, the performance of this classifier was deemed satisfactory. The advantages, limitations and general applicability of the classification step are discussed in greater detail in Section 5.

3.3 Particle Characterization

3.3.1 Primary Crystals

For particles classified as primary crystals, the image analysis routine proceeds to determine the appropriate shape class (spheres, needles, cubuoids, etc.). For particles characterized as needles–as is the case for β L-glutamic acid–the algorithm computes the length, L_1, and width, L_2, of a cylinder that fits the projections. By collecting data from a large number of particles (typically in the order of 10^3 to 10^5), a two-dimensional particle size and shape distribution of the primary crystals can be reconstructed. A more detailed description of the entire procedure, as well as discussions of potential issues have been reported previously,[23,24] and a comparison to more sophisticated reconstruction procedures is presented elsewhere.[38]

3.3.2 Agglomerates

While the strategy used for the quantitative characterization of primary particles has been tested and successfully used in a number of instances,[23–26] no such method was available for the characterization of agglomerates. One of the underlying reasons for this is the wide variety of agglomerate shapes (cf. Figure 9), which is in stark contrast with the lack of available options for the efficient morphological description of agglomerates or aggregates containing non-equant primary particles.[31]

Striving to obtain a computationally simple and robust characterization method while avoiding the above issue, whose solution ultimately lies outside the scope of this work, only a single quantity is extracted from particles classified as agglomerates, namely the agglomerate volume v_a. This measure can be estimated by calculating the volume of the non-convex polytope enclosed by the two projections obtained using our stereoscopic setup. A pragmatic solution to this problem is given by a straightforward 'counting' of voxels. Given two binary images with N_X, N_Y and N_Z pixels along the different axes and a shared z-coordinate, the agglomerate volume of a given particle can be approximated by

Table 2: Overview of operating conditions for the different groups of experiments

Group	Experiments	Seed Fraction	m_s [g]	n_r [rpm]	S_0 [-]
A	1,2,3	small (S)	3	250	5
B	4,5	small (S)	3	400	5
C	6,7	small (S)	3	250	3
D	8,9,10	large (L)	5	250	5
E	11,12	large (L)	5	400	5

$$v_\mathrm{a} \approx \left[\sum_{k=1}^{N_Z}\sum_{i=1}^{N_X} F_{i,k}^{\mathrm{xz}}\left(\sum_{j=1}^{N_Y} F_{j,k}^{\mathrm{yz}}\right)\right] V_\mathrm{voxel} \tag{7}$$

where $F_{i,k}^{\mathrm{xz}}, F_{j,k}^{\mathrm{yz}} \in \{0,1\}$ are the pixel values at a given point in the xz- or the yz-projection, respectively, and $V_\mathrm{voxel} = (1.15\ \mu\mathrm{m})^3$ is the voxel volume. Note that by doing so, an implicit assumption of zero porosity for the occluded parts of the cell is made. Therefore, the computed volume can only be an upper-bound estimate of the true agglomerate volume. Consequently, we choose to visualize population data of agglomerates only in the form of cumulative distributions, which are more readily interpreted in this context.

4 Results

To demonstrate the efficacy and applicability of the proposed supervised learning mechanism for the measurement of agglomeration of needle-like β L-glutamic acid, twelve batch agglomeration experiments were performed under the operating conditions reported in Table 2. The reproducibility of the measurements shall be discussed in detail before we focus our attention on the trends exhibited by the data.

A summary of all experimental results is provided in Table 3, where the number and volume-weighted average lengths and aspect ratios of the primary particles are reported in terms of various moments of the two populations. The definition of the cross-moments for the two-dimensional distribution of primary particles is

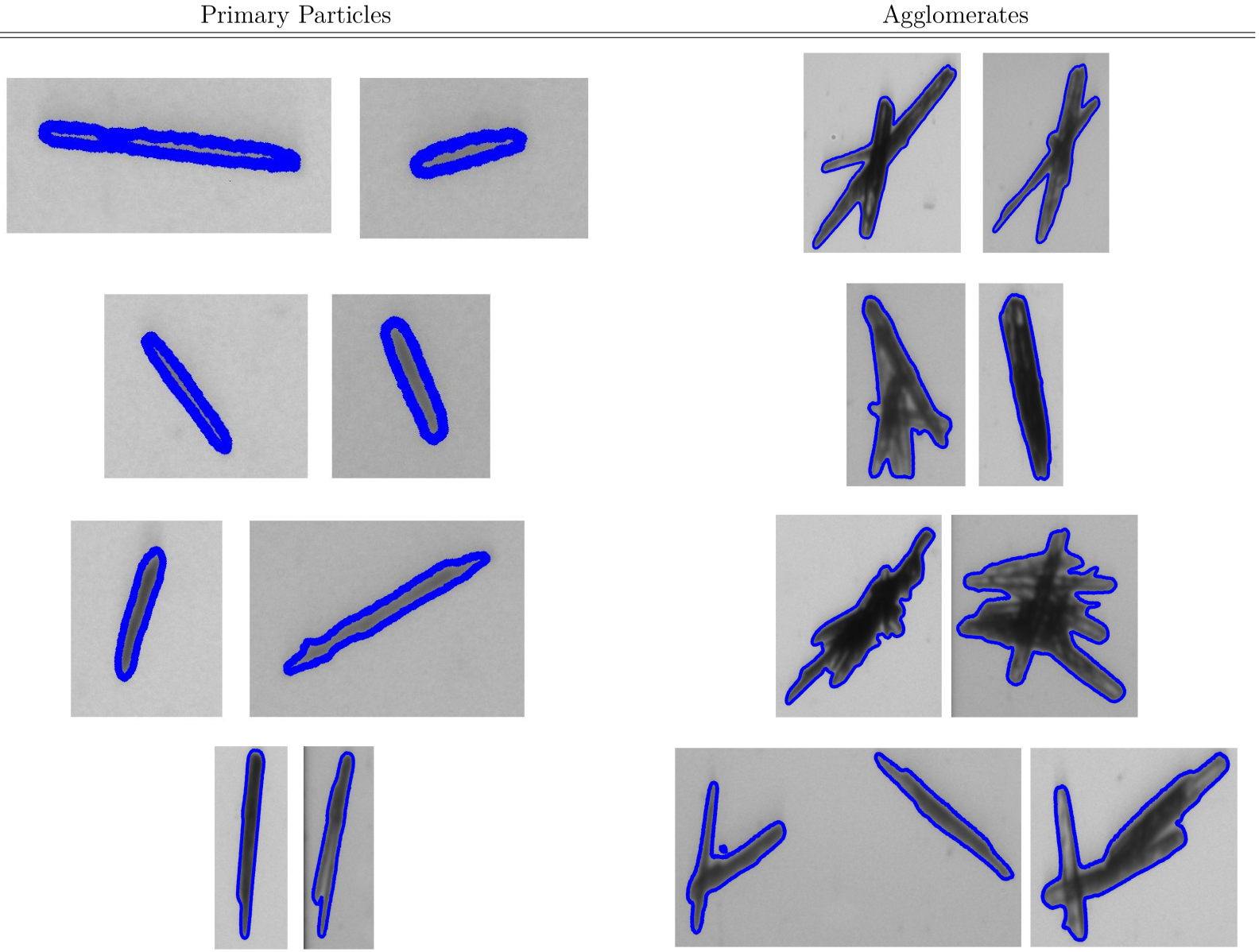

Figure 9: Examples of raw images of primary particles and agglomerates. The blue colour indicates the contour as seen by the computer vision algorithm. Also note that the agglomerate in the lowest row would not be recognized for the analysis because the projections could not be matched between the two cameras (cf Schorsch et al.[23]).

$$\mu_{ij} = \int_0^\infty \int_0^\infty f_\mathrm{s}(L_1, L_2)\ L_1^i\ L_2^j\ \mathrm{d}L_1\ \mathrm{d}L_2 \tag{8}$$

where f_s is the number density function of the primary particles. Similarly, in the case of the agglomerates, the moments are defined as

$$\mu_i^\mathrm{a} = \int_0^\infty f_\mathrm{a}(v_\mathrm{a})\ v_\mathrm{a}^i\ \mathrm{d}v_\mathrm{a} \tag{9}$$

with the corresponding agglomerate number density f_a. In particular, the quantities reported in Table 3 represent: μ_{10}/μ_{00} and μ_{22}/μ_{12} are the number- and volume-weighted average lengths of the primary particles, respectively; μ_{01}/μ_{00} and μ_{13}/μ_{12} are the corresponding average widths; $\mu_{1,\text{-}1}/\mu_{00}$ and μ_{21}/μ_{12} represent the number- and the volume-weighted values of the aspect ratio L_1/L_2, respectively. They are chosen to represent the real, averaged values instead of the alternatives μ_{10}/μ_{01} and μ_{22}/μ_{13}, whose interpretation is much less intuitive. Table 3 further provides information regarding $\mu_{10}^\mathrm{a}/\mu_{00}^\mathrm{a}$, the average volume of the agglomerates and the number and volume weighted agglomeration degrees, defined as

$$X = \frac{N_\mathrm{a,tot}}{N_\mathrm{a,tot} + N_\mathrm{s,tot}} = \frac{\mu_0^\mathrm{a}}{\mu_0^\mathrm{a} + \mu_{00}} \tag{10a}$$

$$X_\mathrm{v} = \frac{V_\mathrm{a,tot}}{V_\mathrm{a,tot} + V_\mathrm{s,tot}} = \frac{\mu_1^\mathrm{a}}{\mu_1^\mathrm{a} + \pi/4\mu_{12}} \tag{10b}$$

As is evident from Table 3, the reproducibility of the volume-weighted properties is–with the exception of the results in group D–satisfactory, however, the number-based measures vary to a much larger extent. A likely cause for this is the occurrence of nucleation in the system, a phenomenon whose presence could not be reliably detected due to fouling of the FBRM probe (cf. Section 2.2.1), yet must be expected to be of some importance at the supersaturation levels reached in this work.[32] Despite our efforts to avoid or reduce nucleation, this interpretation is consistent with the lower, number-based average length,

Table 3: Overview of Seed and Product Properties.

Group	No.	Mean length μ_{10}/μ_{00} [μm]	Mean length μ_{22}/μ_{12} [μm]	Mean width μ_{01}/μ_{00} [μm]	Mean width μ_{13}/μ_{12} [μm]	Mean aspect ratio $\mu_{1,-1}/\mu_{00}$ [-]	Mean aspect ratio μ_{21}/μ_{12} [-]	Mean volume μ_1^a/μ_0^a [μm^3]	Agglom. degree X [-]	Agglom. degree X_v [-]
Seed S	-	223	346	30	53	8.0	7.5	3.7×10^5	26%	34%
Seed L	-	285	597	34	77	8.5	9.1	6.2×10^5	23%	26%
A	1	112	262	28	64	4.1	4.0	5.4×10^5	34%	68%
	2	108	247	28	65	4.0	3.9	4.9×10^5	34%	73%
	3	110	263	29	68	4.0	3.9	6.2×10^5	38%	73%
B	4	264	479	42	79	6.3	6.6	1.3×10^6	48%	62%
	5	238	458	41	83	5.9	6.1	1.0×10^6	52%	63%
C	6	27	470	5	78	5.4	7.1	5.0×10^5	6%	40%
	7	210	499	34	72	6.1	7.6	5.1×10^5	35%	42%
D	8	109	274	28	68	4.0	4.1	8.0×10^5	39%	80%
	9	176	471	43	103	3.9	4.8	2.4×10^6	52%	79%
	10	204	614	45	113	4.5	6.0	1.4×10^6	51%	64%
E	11	260	681	41	108	6.0	7.3	1.9×10^6	38%	57%
	12	12	675	2	111	5.2	7.0	2.0×10^6	3%	61%

average width and agglomeration degree observed in a number of experiments, particularly experiments 6 and 12, and to a smaller extent experiment 8. Due to the higher sensitivity of the number-weighted quantities to this mechanism, we will rely mostly on the interpretation of the volume-weighted quantities in the following.

The normalized, volume-weighted primary particle size and shape distributions (PSSDs) and the normalized, volume-weighted cumulative size distributions of agglomerates for all experiments are plotted in Figure 10 and Figure 11, respectively. Note that the volume-weighted PSSDs were median-filtered prior to plotting in order to reduce artifacts in the contour plots caused by small numbers of very large particles.

The distributions vary significantly for the different conditions and their reproducibility is mostly satisfactory. Regarding group D, careful study of the distributions reveals that experiments 9 and 10 (red and green plots) share a significant similarity, whereas experiment 8 (blue curves) appears to be an outlier. In addition to the clear qualitative difference in the distributions, it is evident both from Figure 10(d) and Table 3 that the average size of primary particles in experiment 8 is much smaller than that of the other two experiments, mostly due to the absence of a great number of large primary particles. Furthermore, experiment 8 contains a much larger number of smaller agglomerates than the other two and exhibits a smaller number-based degree of agglomeration. A possible cause for the observed behavior might be the agglomeration between large particles and newly formed nuclei, which would explain both these observations.

Let us now turn our attention towards some general trends in the data. First, the strong dependence of agglomeration on supersaturation is evident by comparison between group C ($S_0 = 3$) and all other groups ($S_0 = 5$). The final degree of agglomeration for this case is much smaller than in all other cases and the experiments of group C contain the biggest amounts of 'surviving', large primary particles of all experiments that utilize the small seed fraction.

Next, we can analyze the influence of the stirring rate by comparing the results between

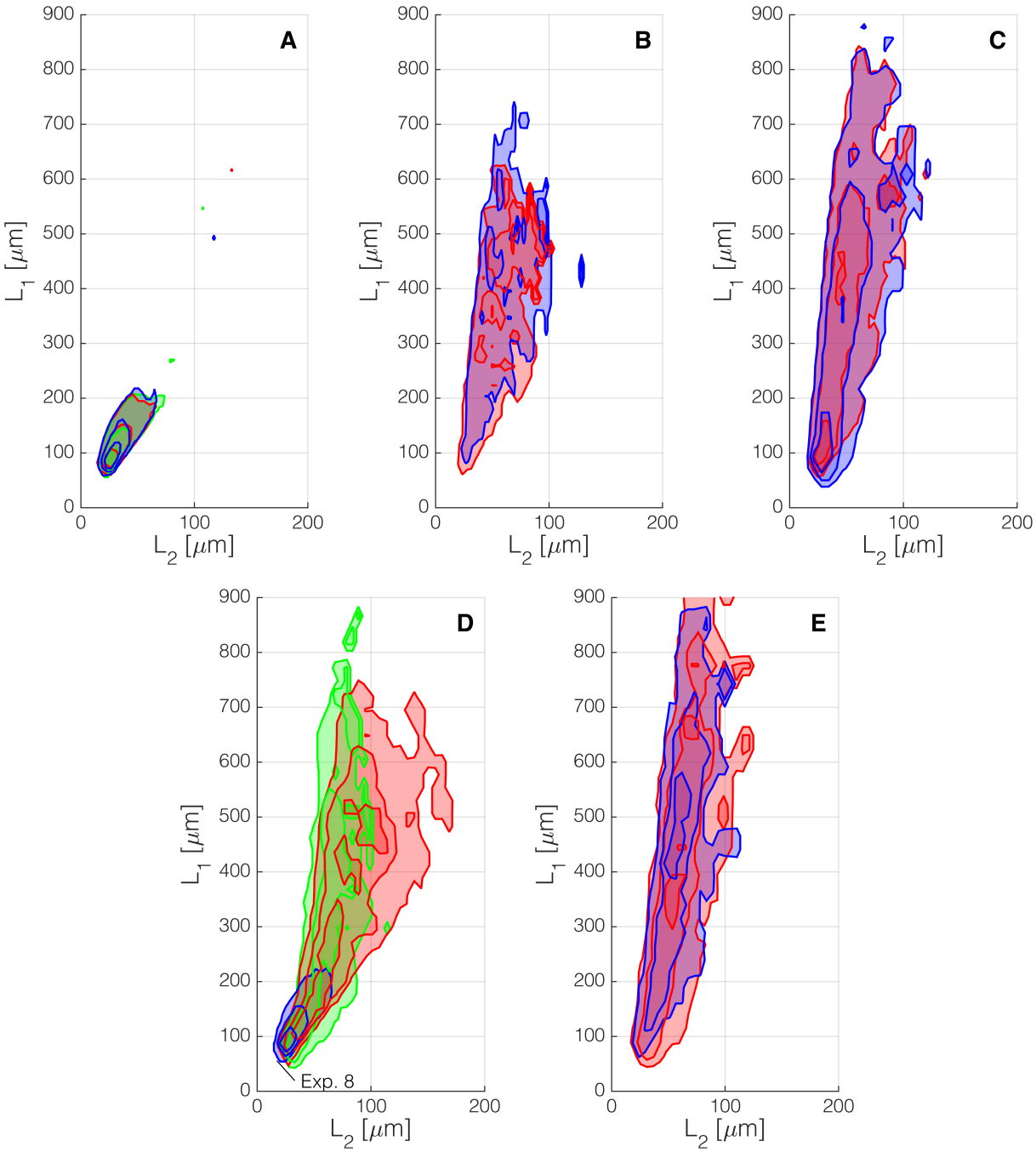

Figure 10: Volume-weighted particle size and shape distributions of experiments in the various groups. Groups A and D contain three experiments, while groups B, C and E contain two (cf. Table 2). All distributions were treated with a median filter prior to plotting in order to reduce artifacts in the contour plots. Different colors (blue, red, green) are used to identify different experiments within the same group. The plotted contour lines represent the 20%, 50% and 80% level sets of a distribution with respect to its maximum.

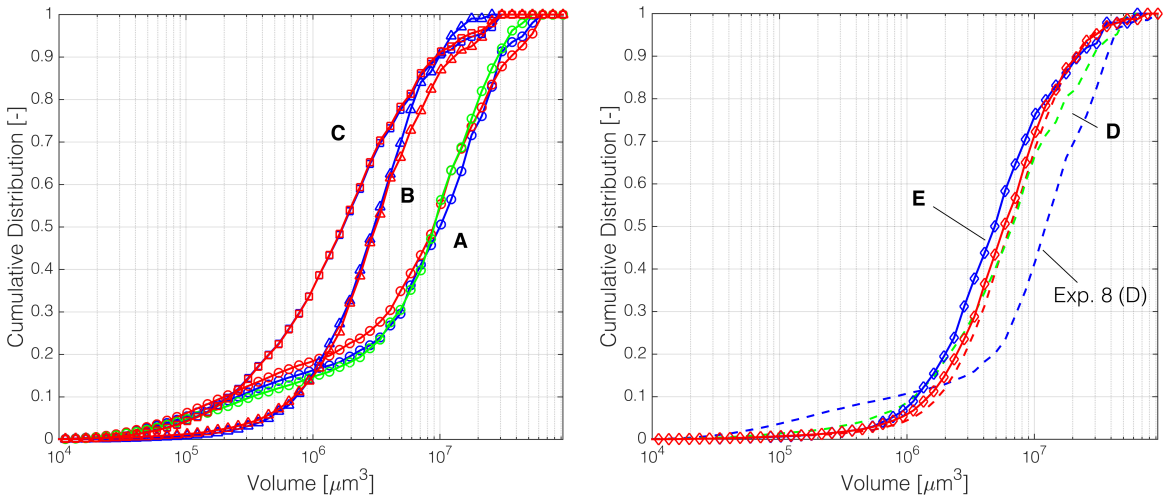

Figure 11: Volume-weighted, cumulative size distribution of agglomerates for groups a) A (circles), B (triangles) and C (squares); b) D (dashed line) and E (diamonds). Shading and different line styles are used to identify different experiments within the same group.

groups A and D ($n_\mathrm{r} = 250$ rpm) and B and E ($n_\mathrm{r} = 400$ rpm), respectively. Both comparisons indicate the agglomeration degree decreases with increasing stirring rate and a greater survival of large primary particles at higher stirring rates, most likely due to the inability of large crystals to form sufficiently strong agglomerate bridges at higher rates.

Furthermore, as to the effect of the seed fraction size, we note that a direct comparison of values is difficult due to different seed masses and seed surface areas. Nevertheless, a comparison of the various distributions clearly shows that the different distributions are dependent on the stirring rate to a much lesser degree in the case of the experiments with large seed fractions (D and E) than in the corresponding experiments with small seed fractions (A and B); most very large primary particles appear unable to agglomerate.

Finally, an interesting observation is that there appears to be a net decrease of the average aspect ratios with increasing agglomeration. A plausible cause for this effect is that higher aspect ratio needles tend to be more likely to agglomerate than those with more equant-sized shapes. This finding, which like all other observations is based on the measurement of tens of thousands of particles, is of particular interest as it implies that a better control of the size and shape (and hence aspect ratio) distribution may allow for an improved ability to avoid

agglomeration. While these findings are preliminary in nature, we believe that they warrant further investigation.

5 Conclusion

In this work, we have presented novel image analysis tools aimed at characterizing an ensemble of particles in terms of two classes, namely that of agglomerates, characterized by a single characteristic size, and that of primary, needle-like crystals, described as cylinders. Attributing measured objects to two classes yields two corresponding particle size distributions, i.e., a 1D PSD of agglomerates and a 2D PSSD of primary crystals. These tools are based on a supervised machine learning strategy and a pragmatic upper bound estimation of the agglomerate volume. After the creation of a suitable training set and demonstrating the ability of the classifier to yield a satisfactory identification performance, the method has been used to investigate the behavior of the β L-glutamic system for a variety of operating conditions. Through the ability to separately study the different distributions, it has been possible to gain considerable insight into the agglomeration of this needle-like substance.

Indeed, we see the capability of distinguishing between different types of crystals based on their morphological features as a critical achievement in the study of crystallization processes. For the case of agglomerated, abraded or broken needles, only morphological features determined through optical tools and image analysis allow for such a classification in a straightforward and reliable manner. It is clear that the multitude of established computer vision strategies and the widespread accessibility of excellent image analysis software have played an important role in profiting of such potential. While we believe that the choice of descriptors plays an important role for the discrimination of objects and that the selection may be further improved by systematic analysis, we attribute the success of our method largely to two specific features of our approach, which are independent of the image analysis itself. Namely, these are the usage of stereoscopic imaging tools on the one hand and the

availability of large data sets on the other. While the former feature allows distinguishing agglomerates from overlapping particle projections in most cases, the latter feature makes the reconstruction of entire distributions with reasonable statistical significance possible. Both contribute significantly and decisively to the robustness of the approach.

We strongly believe in the transferability of the strategy presented in this work to other crystal systems; nevertheless, this remains to be verified and at the same time the performance of the classifier can certainly be further improved upon. However, we believe that it is also important to consolidate the results and the experimental observations reported here by better understanding their underlying principles. To this end, in the next part of this series, novel theoretical models for the agglomeration of non-equant shaped crystals will be presented. The demonstrated ability to measure the different particle types separately will be exploited in order to establish a conceptual link between shape and agglomeration of these crystals. The virtually unexplored relationship of crystal morphology and agglomeration (or breakage) is one important step towards this goal. The mathematical description obtained in this way will be shown to represent a tool for the investigation of such processes in general. Therefore, together with the monitoring and classification tools presented in this first part of the series, it will be able to guide and support future experimental work, as well as to allow for the development of better process design tools and for the ultimate achievement of better product quality.

Acknowledgements

The authors are thankful to the Swiss National Science Foundation for their financial support (project number 200021-155971).

Supporting Information Available

Additional expert agreement analysis and details regarding the machine learning strategies that were tested for the classification of the images. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

Notation

A	area	[m^2]
c	solute concentration in the liquid phase	[kg kg^{-1}]
c^*	solubility	[kg kg^{-1}]
$\mathcal{D}$	Training Set	[-]
$F_{i,j}$	pixel value of image at coordinate i, j	[-]
f	number density function	[varies]
I	image descriptor	[-]
m_s	seed mass	[kg]
N	number of particles	[-]
N_X	number of pixels in X direction	[-]
N_Y	number of pixels in Y direction	[-]
N_Z	number of pixels in Z direction	[-]
n_r	stirring rate	[s^{-1}]
r_i	distance from centroid of particle projection to its boundary	[m]
S	supersaturation	[-]
S_0	supersaturation at time zero	[-]
T	temperature	[K]
t	time	[s]
V	volume	[m^3]
v_a	characteristic agglomerate volume	[m^3]
X	number-based agglomeration degree	[-]
X_v	volume-based agglomeration degree	[-]
x	particle feature vector	[-]
y	classification label	[-]

Greek letters

μ_i^a	moment of agglomerate distribution	[m^{3i} kg^{-1}]
$\mu_{i,j}$	ij-cross moment of shape distribution	[m^{i+j} kg^{-1}]

Sub- and Superscripts

a	variable pertaining to agglomerate distribution	[-]
s	variable pertaining to primary particle distribution	[-]
xz	variable pertaining to xz-projection	[-]
yz	variable pertaining to yz-projection	[-]

References

(1) Limbach, L. K.; Li, Y.; Grass, R. N.; Brunner, T. J.; Hintermann, M. A.; Muller, M.; Gunther,D.; Stark, W. J. Oxide nanoparticle uptake in human lung fibroblasts: effects of particle size, agglomeration, and diffusion at low concentrations *Env. Sci. Technol.* **2005**, *39*, 9370–9376.

(2) Yuk, J. M. and Park, J.; Ercius, P.; Kim, K.; Hellebusch, D. J.; Crommie, M. F.; Lee, J. Y.; Zettl, A.; Alivisatos, A. P. High-Resolution EM of Colloidal Nanocrystal Growth Using Graphene Liquid Cells. *Science* **2012**, *336*, 61–64.

(3) Cameirão, A.; David, R.; Espitalier, F.; Gruy, F. Effect of precipitation conditions on the morphology of strontium molybdate agglomerates. *J. Cryst. Growth* **2008**, *310*, 4152–4162.

(4) Bałdyga, J.; Jasińska, M.; Orciuch, W. Barium Sulphate Agglomeration in a Pipe - An Experimental Study and CFD Modeling. *Chem. Eng. Technol.* **2003**, *26*, 334–340.

(5) Ilievski, D.; White, E. T. Agglomeration during precipitation: Agglomeration mechanism identification for Al(OH)3 crystals in stirred caustic aluminate solutions. *Chem. Eng. Sci.* **1994**, *49*, 3227–3239.

(6) Ilievski, D.; Livk, I. An agglomeration efficiency model for gibbsite precipitation in a turbulently stirred vessel. *Chem. Eng. Sci.* **2006**, *61*, 2010–2022.

(7) Tavare, N.S.; Patwardhan, A. V. Agglomeration in a Continuous MSMPR Crystallizer. *AIChE J.* **1992**, *38*, 377–384.

(8) Zumstein, R. C.; Rousseau, R. W. Agglomeration of copper sulfate pentahydrate crystals within well-mixed crystallizers. *Chem. Eng. Sci.* **1989**, *44*, 2149–2155.

(9) Fujiwara, M.; Chow, P. S.; Ma, D. L.; Braatz, R. D. Paracetamol Crystallization Using Laser Backscattering and ATR-FTIR Spectroscopy: Metastability, Agglomeration, and Control. *Cryst. Growth Des.* **2002**, *2*, 363–370.

(10) Kawashima, Y.; Okumura, M.; Takenaka, H. Spherical Crystallization : Direct Spherical Agglomeration of Salicylic Acid Crystals during Crystallization. *Science* **1982**, *216*, 1127–1128.

(11) Kwon, S.; Nayhouse, M.; Christofides, P. D.; Orkoulas, G. Modeling and control of shape distribution of protein crystal aggregates. *Chem. Eng. Sci.* **2013**, *104*, 484–497.

(12) Lindenberg, L.; Vicum, L.; Mazzotti, M. L -Glutamic Acid Precipitation: Agglomeration Effects. *Cryst. Growth Des.* **2008**, *8*, 224–237.

(13) Yu, Z. Q.; Tan, R. B. H.; Chow, P. S. Effects of operating conditions on agglomeration and habit of paracetamol crystals in anti-solvent crystallization. *J. Cryst. Growth* **2005**, *279*, 477–488.

(14) Mersmann, A. *Crystallization Technology Handbook*. 2nd ed.; Marcel Dekker Inc.; New York, 2001.

(15) Miki, H.; Terashima, T.; Asakuma, Y.; Maeda, K.; Fukui,K. Inclusion of mother liquor inside KDP crystals in a continuous MSMPR crystallizer. *Sep. Purif. Technol.* **2005**, *43*, 71–76.

(16) Mullin, J. W. *Crystallization*, 4th ed.; Butterworth Heinemann: Oxford, U.K., 2001.

(17) Alander, E. M.; Uusi-Penttila, M. S.; Rasmuson, A. C. Characterization of paracetamol agglomerates by image analysis and strength measurement. *Powder Technol.* **2003**, *130*, 298–306.

(18) Faria, N.; Pons, M. N.; Feyo de Azevedo, S.; Rocha, F. A.; Vivier,H. Quantification of the morphology of sucrose crystals by image analysis. *Powder Technol.* **2003**, *133*, 54–67.

(19) Kofler, U.; Riebel, V.; Löffler, F. Einsatz der quantitativen Formanalyse zur Beschreibung von Kristall-Agglomeraten. *Chem.-Ing.-Tech.* **1991**, *63*, 140–142.

(20) Ros, F.; Guillaume,S.; Rabatel, G.; Sevila,F. Recognition of overlapping particles in granular product images using statistics and neural networks. *Food Control* **1995**, *6*, 37–43.

(21) Ma, C. Y.; Wang, X. Z. Model identification of crystal facet growth kinetics in morphological population balance modeling of L-glutamic acid crystallization and experimental validation. *Chem. Eng. Sci.* **2012**, *70*, 22–30.

(22) Ma, C. Y.; Wan, J.; Wang, X. Z. Faceted growth rate estimation of potash alum crystals grown from solution in a hot-stage reactor. *Powder Technol.* **2012**, *227*, 96–103.

(23) Schorsch, S.; Vetter, T.; Mazzotti, M. Measuring multidimensional particle size distributions during crystallization. *Chem. Eng. Sci.* **2012**, *77*, 130–142.

(24) Schorsch, S.; Ochsenbein, D. R.; Vetter, T.; Morari, M.; Mazzotti, M. High accuracy online measurement of multidimensional particle size distributions during crystallization. *Chem. Eng. Sci.*, **2014**, *105*, 155–168.

(25) Ochsenbein, D. R.; Schorsch, S.; Vetter, T.; Mazzotti, M.; Morari,M. Growth Rate Estimation of β-L-Glutamic Acid from online measurements of multidimensional particle size distributions and concentration. *Ind. Eng. Chem. Res.* **2014**, *53*, 9136–9148.

(26) Ochsenbein, D. R.; Schorsch, S.; Salvatori, F.; Vetter, T.; Morari,M.; Mazzotti, M. Modeling the Facet Growth Rate Dispersion of β L-Glutamic Acid – Combining Single Crystal Experiments with nD Particle Size Distribution Data. *Chem. Eng. Sci.* **2015**, submitted.

(27) Meakin, P. Collisions between Point Masses and Fractal Aggregates. *Langmuir* **1989**, *5*, 510–518.

(28) Xiong, Y.; Pratsinis, S. E. Formation of agglomerate particles by coagulation and sintering - Part I. A two-dimensional solution of the population balance equation. *J. Aerosol Sci.* **1993**, *24*, 283-300.

(29) Hashino, T.; Tripoli, G. J. The Spectral Ice Habit Prediction System (SHIPS). Part III: Description of the Ice Particle Model and the Habit-Dependent Aggregation Model. *J. Atmos. Sci.* **2011**, *68*, 1125–1141.

(30) Borchert, C.; Sundmacher, K. Crystal Aggregation in a Flow Tube: Image-Based Observation. *Chem. Eng. Technol.* **2011**, *34*, 545–556.

(31) Briesen, H. Hierarchical Characterization of Aggregates for Monte Carlo Simulations. *AIChE J.* **2006**, *52*, 2436–2446.

(32) Cornel, J.; Lindenberg, C.; Mazzotti, M. Experimental Characterization and Population Balance Modeling of the Polymorph Transformation of L-Glutamic Acid. *Cryst. Growth Des.* **2009**, *9*, 243–252.

(33) Schöll, J.; Bonalumi, D.; Vicum, L.; Mazzotti, M. In Situ Monitoring and Modeling of the Solvent-Mediated Polymorphic Transformation of L -Glutamic Acid. *Cryst. Growth Des.* **2006**, *6*, 881–891.

(34) ten Cate, A.; Derksen, J. J.; Portela, L. M.; Van den Akker, H. E. A. Fully resolved simulations of colliding monodisperse spheres in forced isotropic turbulence. *J. Fluid Mech.* **2004**, *519*, 233–271.

(35) Marchisio, D. L.; Soos, M.; Sefcik, J.; Morbidelli, M. Role of Turbulent Shear Rate Distribution in Aggregation and Breakage Processes. *AIChE J.* **2006**, *52*, 158–173.

(36) Bouwman, A.; Bosma, J.; Vonk, P.; Wesselingh, J.; Frijlink, H. Which shape factor(s) best describe granules? *Powder Technol.* **2004**, *146*, 66–72.

(37) Hofmann, T.; Schöllkopf, B.; Smola, A.J. Kernel Methods in Machine Learning. *Ann. Stat.* **2008**, *36*, 1171–1220.

(38) Schorsch, S.; Hours, J.-H.; Vetter, T.; Mazzotti, M.; Jones, C. N. An optimization-based approach to extract faceted crystal shapes from stereoscopic images. *Comput. Chem. Eng.*, **2015**, *75*, 171–183.

For Table of Contents Use Only

Title: Agglomeration of needle-like Crystals in Suspension – I. Measurements

Authors: David R. Ochsenbein, Thomas Vetter, Stefan Schorsch, Manfred Morari and Marco Mazzotti

Synopsis: Image analysis is used to detect agglomerates in a crystallization process. Subsequently, the size of agglomerates and the size and shape of primary, needle-like particles are computed, allowing for the reconstruction of two separate particle size distributions for the two classes. The detailed information is then used to study the agglomeration behavior of a specific system under varying operating conditions.