Statistical Analysis and Nucleation Parameter Estimation from Nucleation Experiments in Flowing Microdroplets

Elena Cândida dos Santos, † Giovanni Maria Maggioni, †§ and Marco Mazzotti ‡, †©

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

ABSTRACT: We have studied the primary nucleation of adipic acid from aqueous solutions in thousands of microdroplets generated in a fully automated microfluidic setup. By varying supersaturation in solution and residence time, we were able to estimate nucleation rates and growth times, while accounting for the stochastic nature of nucleation, the variability in microdroplet volumes (which is kept below 2%, thanks to a carefully designed experimental protocol), and the uncertainty in the automated image analysis procedure. Through a thorough statistical analysis we have obtained exact expressions for the expected values and the variances of all the random variables involved, all the way to the nucleation rate and the growth time associated with each supersaturation level explored and to the model parameters appearing in the corresponding constitutive equations. We have analyzed what controls the overall uncertainty in the estimation of the physical quantities above. We have shown that the distribution of droplet volumes at the level observed here is not limiting, whereas the variability in the experimental accuracy and of error propagation may still be a concern. Such an experimental protocol is in principle simple and appealing. However, it is not devoid of issues, both experimental and theoretical, which make a careful consideration of experimental accuracy and of error propagation necessary in order to be able to attain estimates not only of nucleation rates and related model parameters but also of their uncertainty. Sources of uncertainty are several, namely the control of the experimental conditions themselves, the possibility of droplet volumes to be distributed, and the intrinsic errors associated with the observation of droplets and the detection of crystals therein, as well as the fact that there are multiple dependences that have to be explored experimentally in a statistically relevant manner, namely that of the nucleation probability on residence time and droplet volume and that of the nucleation rate on supersaturation and temperature.

The purpose of this work is a thorough analysis of these issues, based on a comprehensive set of experimental data on the nucleation of adipic acid from aqueous solutions in microdroplets of about 170 nL. The manuscript is divided into three parts, where the experimental setup and procedure are presented first (section 2), followed by the theoretical analysis of the statistics describing the experiments, accounting for all types of uncertainties mentioned above (section 3). Finally, in section 4 the theory is applied to our experiments, before finally conclusions are drawn.

1. INTRODUCTION

Since nucleation is stochastic, even under exactly the same experimental conditions the time for nucleation from a supersaturated solution to occur and for new crystals to be observed can vary a great deal. In order to estimate the nucleation rate, i.e., a quantity of great theoretical and practical importance, proper statistics need to be constructed, which entail a number of experiments and measurements as large as possible. Microfluidic platforms, which can generate large numbers of nearly identical microdroplets that can serve as crystallization microenvironments, have been proposed and utilized to build such statistics.1–3 Although other protocols have been proposed, a typical experiment involves observing identical droplets that have had a predetermined time to form and grow new crystals;4–6 the fraction of crystal-containing droplets estimates the nucleation probability (the larger the number of droplets, the more accurate the estimate), which can then be correlated to the nucleation rate. Repeating experiments under different conditions allows determining nucleation rates as a function of, for example, supersaturation.

Such an experimental protocol is in principle simple and appealing. However, it is not devoid of issues, both experimental and theoretical, which make a careful consideration of experimental accuracy and of error propagation necessary in order to be able to attain estimates not only of nucleation rates and related model parameters but also of their uncertainty. Sources of uncertainty are several, namely the control of the experimental conditions themselves, the variability in microdroplet volumes (which is kept below 2%, thanks to a carefully designed experimental protocol), and the uncertainty in the automated image analysis procedure. Through a thorough statistical analysis we have obtained exact expressions for the expected values and the variances of all the random variables involved, all the way to the nucleation rate and the growth time associated with each supersaturation level explored and to the model parameters appearing in the corresponding constitutive equations. We have analyzed what controls the overall uncertainty in the estimation of the physical quantities above. We have shown that the distribution of droplet volumes at the level observed here is not limiting, whereas the variability in the experimental accuracy and of error propagation may still be a concern. Such an experimental protocol is in principle simple and appealing. However, it is not devoid of issues, both experimental and theoretical, which make a careful consideration of experimental accuracy and of error propagation necessary in order to be able to attain estimates not only of nucleation rates and related model parameters but also of their uncertainty. Sources of uncertainty are several, namely the control of the experimental conditions themselves, the possibility of droplet volumes to be distributed, and the intrinsic errors associated with the observation of droplets and the detection of crystals therein, as well as the fact that there are multiple dependences that have to be explored experimentally in a statistically relevant manner, namely that of the nucleation probability on residence time and droplet volume and that of the nucleation rate on supersaturation and temperature.

The purpose of this work is a thorough analysis of these issues, based on a comprehensive set of experimental data on the nucleation of adipic acid from aqueous solutions in microdroplets of about 170 nL. The manuscript is divided into three parts, where the experimental setup and procedure are presented first (section 2), followed by the theoretical analysis of the statistics describing the experiments, accounting for all types of uncertainties mentioned above (section 3). Finally, in section 4 the theory is applied to our experiments, before finally conclusions are drawn.

2. EXPERIMENTAL SECTION

2.1. Experimental Setup. The main purpose of the experimental setup is to create a segmented flow, where microdroplets are continuously generated to act as individual (micro)crystallizers to study primary nucleation. A T-junction followed by 4 m long FEP tubing is used as the microfluidic platform. The stochastic behavior of
nucleation is enhanced in small volumes; hence, the method applied in this work allows sampling of a large amount of data (droplets are mutually independent) to estimate kinetic parameters of nucleation. A similar setup was already presented by Rossi et al. but was operated under different fluid-dynamic conditions and at another droplet size scale. Figure 1 shows the schematic of the experimental setup, which consists of four zones; each of them serves a specific purpose and will be explained in detail in the following.

2.1.1. Generation Zone. First, in the generation zone (GZ), as the name suggests, droplets are generated by feeding the immiscible, dispersed, and continuous fluids into a PEEK (polyether ether ketone) T-junction (Upchurch Scientific, 0.020 in. through hole) using two precision syringe pumps (Harvard Apparatus, Pico Elite Plus). Both phases, prior to meeting at the T-junction, are continuously filtered with hydrophilic filters (Sartorius, Minisart NML Plus glass fiber 0.2 μm surfactant free cellulose acetate) to ensure a constant and low concentration of foreign particles in solution. On the continuous fluid line, a miniature low-pressure flow-through sensor (26PC Series, range 0−5 psi, Honeywell) is installed to monitor the pressure drop along the tube. For a more accurate measurement, the device is calibrated against a sensor of higher accuracy (range 0−2 bar, accuracy ±4 mbar, Omega). After formation, droplets are flown into a translucent FEP (fluorinated ethylene propylene) tubing (Upchurch Scientific, 1/16 in. OD × 0.020 in. ID, 508 μm nominal inner diameter), inserted in an in-house-built near-infrared (NIR) sensor for droplet frequency measurements. The NIR sensor is built based on the design presented by Ladosz et al., with the modification of having inserted only one pair of diodes and detector perpendicular to the tube. The NIR sensor consisted of a cube made of PEEK (non-transparent) with a through hole of 1.6 mm diameter, into which the FEP tubing was inserted. Perpendicular to this, another opening (diameter gradually reduced to 0.5 mm, coinciding with the tube ID) was bored to install, on one side, a diode emitting light in the near-infrared region with a peak wavelength at 940 nm (TSAL4400, Vishay Semiconductors) and, on the other side, the corresponding IR

Figure 1. Fully automated droplet-based microfluidic crystallization setup.

Figure 2. Droplet volume and frequency measurement. (A) NIR absorbance at a wavelength of 940 nm monitoring over time. Insert: example of NIR response for silicone oil−aqueous solution flow. (B) Droplet volume monitoring over time. (C) Pressure drop along the total length of the tube over time. (D) Histogram and corresponding gamma distribution (red line) of droplet volumes.
detector (TSL261-RLF, ams), which was further combined with an impedance amplifier.

The characteristic intensity of each media passing was detected by the NIR sensor, through a combination of reflection and refraction effects, thus enabling their detection. Figure 2A exemplifies the signal output for a two-phase segmented flow monitored over time. In the inset of the same figure, peaks correspond to the dispersed phase (droplets) and the baseline to the continuous phase, which exhibit a repeated pattern that resembles a periodic behavior.

There are two pieces of quantitative information that can be extracted from such signal: first, the distance between the inception of successive droplets, i.e., \( \Delta t_{fi} \), and second, the total number of droplets in an experiment, \( n \). Assuming a constant flow rate of the dispersed phase, \( q_{t} \), over the entire duration of the experiment, \( t_{exp} \), and enforcing the condition that the amount of the dispersed phase injected is distributed within all the \( n \) droplets generated yield the following estimate of the mean droplet volume, \( \mu_{v} \):

\[
\mu_{v} = \frac{q_{t} \times t_{exp}}{n}
\]

(1)

We attribute the variability of the measured quantity \( \Delta t_{fi} \) to an intrinsic variability in how droplets discharge from the T-junction; hence, we consider \( \Delta t_{fi} \) a random variable, with a mean and variance. From the \( \Delta t_{fi} \) value, the droplet volume \( v_{i} \) can be calculated by assuming that the dispersed phase flow rate, \( q_{d} \), and that of the continuous phase, \( q_{c} \), remain constant during the whole experiment (hence the volumetric proportion of the two phases is the same in all fluid plugs between \( t_{fi-1} \) and \( t_{fi} \) for all \( k \) values; such assumption is confirmed by the regularity of the pressure drop profile during the course of the whole experiment, as shown in Figure 2C):

\[
v_{i} = q_{d} \times \Delta t_{fi}
\]

(2)

This procedure applied to the total number of droplets yields a droplet volume distribution, \( \psi (v) \), here described by the gamma distribution, characterized by mean value \( \mu_{v} \), variance \( \sigma_{v}^{2} \), and coefficient of variation \( \psi = \sigma_{v}/\mu_{v} \). For the sake of clarity, in Figure 2B we plot droplet volumes recorded for 30 min. During this time, with a mean frequency of 0.5 Hz, we could produce 900 droplets with \( \mu_{v} = 167 \) nL and \( \psi = 1.88% \). In Figure 2D the same droplet volume distribution in the form of a histogram and its regressed gamma distribution are plotted. Note that the histogram of droplet volumes and the corresponding regressed gamma distribution are not the same quantity, and hence, they do not necessarily overlap; in fact, the distribution is normalized, the histogram is not. The comparison is made to show that the two span the same range of droplet volumes and have the same mode.

As shown in Figure 1, syringe pumps, the T-junction, and the NIR sensor were placed inside a temperature-controlled incubator, always kept at 40 °C, in order to maintain solutions at the desired dissolved state and to stay below the maximum operating temperature of the syringe pumps (45 °C).

2.1.2. Crystallization Zone. After exiting the GZ, droplets flow to the crystallization zone (CZ), which consists of a FEP tubing of a specific length, placed in a water bath at 12.6 °C. The temperature of the water bath is internally controlled using a thermostat (Huber Pilot ONE, Offenburg) and measured with a pt100 sensor to guarantee constant crystallization temperatures. Due to the supersaturation created by cooling, a number of droplets crystallize while traveling along the tube; the tube length can be varied in order to vary the residence time.

We have verified that, under the experimental conditions adopted, droplets entering the CZ need to travel less than 10 mm to cool from 40 to 12.6 °C; at their average velocity, i.e., 2.06 mm/s, this corresponds to a cooling rate of about 7.5 °C/s. Since the tube lengths used in the experiments were on the order of 1 m or more, droplets can be assumed to attain the crystallization temperature instantaneously when they enter the CZ.

Note that temperature variations in the CZ are within the precision limits of the PT100 sensor, i.e., ±0.2 °C, which represents a ±1% variation of the equilibrium concentration. Considering also the precision of the balance (0.001 g) yields an uncertainty in supersaturation of about 1.5%.

2.1.3. Quench Zone. When droplets exit the CZ, they enter the quench zone (QZ) (the tube length inside this zone is 19 cm long, to be on the safe side), which is kept at a temperature 2 °C lower than that corresponding to a saturated solution at the feed concentration.

Under these conditions, there is no further nucleation and negligible growth, but crystals that have already formed cannot dissolve. Moreover, we have verified that the quenching temperature is correctly chosen for each solution concentration, by setting the CZ temperature equal to that of the QZ and verifying that no nucleation occurred even for long residence times.

2.1.4. Observation Zone. Finally, droplets move into the observation zone (OZ), where the FEP tube is placed under a light microscope (Zeiss, AxiosPlan), with which the nature of the segmented flow is observed. An AxiosCam MRc5 is used to record the videos, containing the information on whether droplets carry crystals or not. The camera operates at a frame rate of roughly 12 fps; this together with the neonlue 10X objective, at a binning 8 × 8, yields images of the size 320 × 240 pixels and a resolution of 2.12 µm/pixel. For exemplary images of droplets containing crystals, selected from a recorded video, refer to Figure S1 in the Supporting Information.

2. Chemicals. Adipic acid (hexanedioic acid, (CH₂)₄(COOH)₂, purity ≥99%, purchased from Sigma-Aldrich, U.K.) was used as received, and solutions were made in ultrapure deionized water (Millipore, 18.2 MΩ cm). The solubility of the only known form of crystals of adipic acid in water was determined by interpolating experimental data, covering the temperature range 0–40 °C, reported in the literature and given as a function of temperature by the relationship:

\[
e^{\psi}(T) = 7.72 \times 10^{4} \exp \left( -4.46 \times 10^{-2} T \right)
\]

(3)

where, \( e^{\psi} \) and the pre-exponential factor are in grams of solute per grams of solvent (g/g), while T and the factor inside the exponent are in K. Silicone oil (Huber, M40.165.10, linear polydimethylsiloxane (PDMS)) was selected as the continuous fluid to form the segmented flow, because it features high wettability of the channel walls, insignificant partition with adipic acid, and moderate viscosity (ability to be filtered while flowing).

2.3. Experimental Procedure. Nucleation experiments in droplets were conducted using silicone oil and an aqueous solution of adipic acid, as the continuous and dispersed phases, respectively. Solution quantities were weighed in a balance (Mettler Toledo, precision 0.005 g). Around 40 g of particle-free, deionized water and the desired amount of adipic acid were filled in a beaker, covered with Parafilm to minimize evaporation losses, and stirred at 40 °C overnight, thus allowing enough time for dissolution. Various initial concentrations of the adipic acid solutions were prepared to enable investigating different supersaturations: i.e., nominal supersaturations of 2.08, 2.18, 2.28, and 2.37 were created at the chosen CZ temperature. Note that supersaturation has always been computed using the value of solubility at 12.6 °C; i.e., 0.0128 g/g.11 Before the experiment was started, the temperature of the CZ and QZ was set to 40 °C, and 10 mL of water was flushed to dissolve any adipic acid crystals that might have remained in the tube. The total tube length was fixed to 4 m (total volume equal to 0.81 mL), so that the pressure drop, \( \Delta P \), along the tube was constant for all experiments (see Figure 2C for an illustration). Then, the tube was flushed with silicone oil (\( q_{t} = 200 \) µL/min), thus wetting the tube entirely to guarantee proper formation and downstream flow of liquid droplets. Subsequently, both flow rates were set to 100 µL/min for a few minutes to get rid of air bubbles and soon thereafter set to the experimental values reported in Table 1.

The experiments in this work were carried out at very low values of the capillary number, \( C_{a} \), and of the Reynolds number, Re. It is known that in such regimes the lubricating film around the droplet is
thin, thus enabling the approximation of the droplet velocity to the superficial velocity, \( U = Q/A \) (where \( Q \) is the total flow rate and \( A \) is the tube cross section). Accordingly, the residence time (i.e., how long droplets travel at the crystallization temperature) is reliably calculated by the ratio of the tube length placed inside the CZ and the droplet velocity: i.e., \( \tau = L_{\text{tube}}/U \). Thus, after placing a certain known length of the FEP tubing was placed inside the CZ, the temperatures of CZ and QZ were set to the experimental values. The CZ takes 40 min to cool to 12.6 °C (temperature measured by the Pt100 sensor inside the water bath), whereas the system needs 1 h to equilibrate the CZ and QZ temperatures, were kept the same. A summary of the operating conditions is provided in Table 1.

### 2.4. Crystal Detection Procedure

The images of the droplets traveling through the observation zone recorded on video can be evaluated by both a human operator and an automated image analysis algorithm. The goal is to determine which droplets contain crystals and which do not. To this end, we processed the movies by applying foreground detection, using the Computer Vision Toolbox of Matlab, which highlights the silhouette of the traveling droplet. The human operators based their assessment on visual observation. The automated procedure exploits a machine learning routine in Matlab, on the basis of a decision tree algorithm, chosen due to its computational efficiency and its capability of performing multiclass classification on a data set.\(^{12}\)

### 3. THEORY

#### 3.1. Statistics of Nucleation Experiments in Identical Microdroplets

A single nucleation experiment in microdroplets consists of the generation of \( n \) virtually identical droplets, having a volume \( v \) and containing a solution that at the selected temperature and initial concentration exhibits a supersaturation \( S \). All droplets are observed at the same residence time \( \tau \), when it is determined whether they contain a crystal (i.e., nucleation has occurred and particles have grown enough to be observable) or not (i.e., either nucleation has not occurred, or it has occurred but nuclei are too small to be observed). Due to the experimental protocol utilized droplets are isolated from each other, hence each droplet represents an independent crystallizer.

From a statistical point of view, the outcome of the observation of each droplet (success, i.e., yes, when crystals are present, and failure, i.e., no, when they are not) constitutes a Bernoulli trial. The success probability, \( p \), depends on the three operating conditions above, i.e., \( v \), \( S \), and \( \tau \), and it is in principle the same for all droplets, since droplets are ideally identical. As such, the sequence of outcomes from the observation of the \( n \) droplets belonging to one experiment constitutes a Bernoulli process, whose outcome in a specific realization is the number of successes, thus the random variable \( h \), with \( 0 \leq h \leq n \). The random variable \( h \) follows the binomial distribution that governs Bernoulli processes, i.e., the probability that \( h \) is exactly \( k \) is

\[
\Pr(h = k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

Therefore, the expected value of \( h \) is \( E[h] = np \) and its variance is \( \text{Var}[h] = np(1 - p) \).

It is worth noting that the success probability under the operating conditions selected is indeed the quantity of physical interest. There is a one-to-one mapping between the success probability \( p \), estimated through a nucleation experiment in microdroplets, at specific values of \( v \), \( S \), and \( \tau \), and the nucleation rate, \( j \), at the same level of supersaturation. Such correspondence is based on the observation that each microdroplet is a (micro)crystallizer, where stochastic nucleation is assumed to be a Poisson process. Therefore, the nucleation time in each individual droplet is distributed according to the relevant exponential distribution, whose process intensity is the product of crystallizer volume, \( v \), and nucleation rate, \( j \),\(^{13-15}\);

As a consequence, when an ensemble of identical droplets is considered, the value of the success probability, \( p \), represents an estimate also of the exponential cumulative probability corresponding to the specific value of residence time, \( \tau \), selected (this is the probability that nucleation occurs in a droplet at any time before \( \tau \)). By carrying out experiments with the same \( v \) and \( S \) but at different \( \tau \), one can obtain estimates of the cumulative probability at different residence times and hence an estimate of the whole cumulative probability distribution. Such distribution for a Poisson process can be written as

\[
F = 1 - \exp(-jv(\tau - \tau_g))
\]

where \( \Delta \tau = \tau - \tau_g \) is the nucleation time, \( \tau \) is the detection time that corresponds to the experimental residence time, and \( \tau_g \) is the growth time, which is necessary for nuclei to grow to a detectable size, depends on the experimental supersaturation as well, and needs to be estimated from the experiments.\(^{16}\) On the basis of the discussion above, the cumulative probability \( F \) and the success probability \( p \) represent the same physical quantity.

#### 3.2. Droplet Volumes and Success Probabilities Are Distributed

In a real experiment, however, droplets are not identical and droplet volumes are distributed, as shown in Figure 2 and discussed in section 2.1.1. Since at constant supersaturation and residence time the success probability and the droplet volume are related one to one through eq 5, with \( F = p \), success probabilities are also distributed. The normalized volume distribution is called \( \phi(v) \) and is characterized in this work as a gamma distribution with parameters \( \alpha \) and \( \beta \):

\[
\phi(v) = \frac{\beta^\alpha v^{\alpha-1} e^{-\beta v}}{\Gamma(\alpha)}
\]

where \( \Gamma(\cdot) \) is the gamma function. Therefore, the expected value of \( v \) is \( E[v] = \mu_v = \alpha/\beta \), its variance is \( \text{Var}[v] = \sigma_v^2 = \alpha/\beta^2 \), and its coefficient of variation is \( 1/\sqrt{\alpha} \). Note also that our experimental protocol allows us to obtain droplets with a coefficient of variation as small as 2% (see Figure 2). Even though variations up to 15% have been reported in the

| fluid              | \( Q \) (µL/min) | \( U \) (mm/s) | \( f \) (Hz) | \( \mu_c \) (nL) |
|--------------------|------------------|--------------|-------------|-----------------|
| adipic acid (aq)   | 5                | 2.06         | 0.5         | 167             |
| silicone oil       | 20               |              |             |                 |

Table 1. Experimental Operating Conditions

---

Cryst. Growth & Des. Article

DOI: 10.1021acs.cgd.9b00562

Cryot. Growth Des. 2019, 19, 6159–6174

---
literature and considered acceptable, we will question this consideration on the basis of the results reported below.

The corresponding normalized distribution of success probabilities, \( \pi(p) \), is calculated by enforcing the condition \( \phi(v) \ dv = \pi(p) \ dp \) and using eq 5, with \( F = p \). One obtains

\[
\pi(p) = \frac{1}{fJ\Delta r(1 - p)} \phi\left(1 + \frac{1}{fJ\Delta r} \ln \frac{1}{1 - p}\right) \tag{7}
\]

The expected value and the variance of \( \pi(p) \) are called \( \mu_p \) and \( \sigma_p^2 \), respectively, and are calculated as

\[
\mu_p = \int_0^1 p\pi(p) \ dp = 1 - \left(1 + \frac{\mu J\Delta r}{\alpha}\right)^{-\alpha} \tag{8}
\]

\[
\sigma_p^2 = \int_0^1 (p - \mu_p)^2\pi(p) \ dp = \left(1 + \frac{2\mu J\Delta r}{\alpha}\right)^{-\alpha} - \left(1 + \frac{\mu J\Delta r}{\alpha}\right)^{-2\alpha} \tag{9}
\]

Note that these two quantities converge correctly to \( (1 - \exp(-\mu J\Delta r)) \) and to zero, respectively, when \( \alpha \to \infty \) (or else the coefficient of variation of the droplet volume distribution approaches 0): i.e., when the droplet volume distribution approaches a Dirac delta centered in \( \mu_p \). It is also worth noting that when the coefficient of variation increases from 2% to 15%, while everything else remains constant, the value of \( \mu_p \) varies of about 0.5%, whereas the variance \( \sigma_p^2 \) becomes more than 50 times larger.

The sequence of outcomes from the observation of nucleation in \( n \) droplets with nucleation probabilities distributed as in eq \( 7 \) is a random variable \( h_p \), which is the outcome of the sum of \( n \) independent nonidentical Bernoulli trials. It follows a "generalized" binomial distribution, i.e., the probability that \( h_p = k \) is

\[
\Pr(h_p = k) = \int_0^1 \binom{n}{k} \frac{\mu}{k}^k (1 - \frac{\mu}{k})^{n-k} \pi(p) \ dp \tag{10}
\]

The expected value of \( h_p \) is \( n\mu_p \) since combining eq 10 and eq 8 yields

\[
E[h_p] = \sum_{k=0}^n k\Pr(h_p = k) = \int_0^1 np \sum_{k=1}^n \binom{n-1}{k-1} \frac{\mu}{k}^{k-1} (1 - \frac{\mu}{k})^{n-k} \pi(p) \ dp = n\mu_p \tag{11}
\]

Its variance is given by the expression

\[
V[h_p] = n\mu_p(1 - \mu_p) + n(n - 1)\sigma_p^2 \tag{12}
\]

where we have used the identities \( V[h_p] = E[h_p^2] - (E[h_p])^2 \) and \( V[p] = \sigma_p^2 = E[p^2] - \mu_p^2 \) as well as the following relationship for \( E[h_p^2] \) (obtained by combining again eqs 10 and 8 and the last two identities):

\[
E[h_p^2] = \sum_{k=0}^n k^2\Pr(h_p = k) = \sum_{k=0}^n (k + k(k - 1))\Pr(h_p = k)
= n\mu_p + n(n - 1)E[p^2] \tag{13}
\]

It is apparent from eq 12 that the variance of \( h_p \) is the sum of the variance of the outcome of a Bernoulli process with success probability equal to \( \mu_p \) and the variance of the distribution of the success probability multiplied by a factor approximately equal to the square of the number of Bernoulli trials (i.e., of droplets). Since we are going to estimate the mean of the distribution of the success probabilities as \( h_p/n \), we can use eq 12 to calculate the variance of such an estimate:

\[
V\left[\frac{h_p}{n}\right] = \frac{V[h_p]}{n^2} = \frac{\mu_p(1 - \mu_p)}{n} + \frac{(n - 1)\sigma_p^2}{n} \tag{14}
\]

It is worth noting that the Bernoulli process contribution is inversely proportional to the number of trials, i.e., of droplets (which in our experiments is, with two exceptions, between about 1000 and 3000), whereas the contribution due to the droplet volume distribution is virtually independent of the number of droplets and equal to the variance of the \( \pi(p) \) distribution of success probabilities.

The dependence on \( \mu_p \) of the left-hand side of the last equation (blue line) and of the two terms on the right-hand side (red and green lines for the first and second parts, respectively) is illustrated in Figure 3. The solid lines (in the main figure and in the inset) are characteristic of the data in this work: i.e., \( n = 1000 \) (red curve) and \( \sigma_p^2 \) corresponding to a coefficient of variation in the distribution of volumes of 2%. The dashed lines in the main figure are obtained for \( \psi = 15\% \). The blue curves, the left-hand side of eq 14, represent the overall variance as a function of \( \mu_p \).
dominated by the intrinsic variance of the Bernoulli process (which applies also when all droplets are identical) in our case, where we painstakingly control the narrowness of the volume distribution, whereas it is dominated by the variance in the volume distribution for the data reported in the literature above. The maximum value of the variance in the latter case is about 10 times larger than in the former.

In the context of microfluidic studies of nucleation, this analysis points at the importance of measuring and reporting the droplet volume distribution in detail and in controlling it tightly, so as to minimize its effect on the propagation of experimental uncertainties. It is worth noting that, in our previous study on droplet volume distributions, we had concluded that even a 15% value of the coefficient of variation of the volume distribution would be acceptable for an accurate estimate of the nucleation rate through experiments such as those carried out in this study. That consideration is in principle valid if one considers that the mean value of the success probability is not affected by the coefficient of variation of the volume distribution strongly. However, the variance of, and hence the uncertainty on, the mean value of the success probability increases quickly when such a coefficient of variation increases (see eqs 8 and 9 and the exemplary values given in the discussion thereof).

3.3. Solid Detection in Droplets Is Error Prone. In the previous sections, we considered the outcome of the observation of each droplet in an experiment consisting of $n$ droplets as a success when crystals are present and as a failure vice versa. In practice, the online camera (see section 2.4) captures images of all $n$ droplets; out of these $n$ images, $n_\text{a}$ images (about 10% of the $n$ droplets in an experiment) are evaluated both by a human operator and by an automatic image analysis software (which ultimately analyzes all images for all $n$ droplets in each experiment). Neither is perfect; hence, the number of successes determined in an experiment, i.e., $h_\text{d}$ where the subscript refers to detection, is itself a random variable. Its realization in a specific experiment depends on the statistics of the Bernoulli process and of the specific droplet volume distribution, as well as on the statistics of the image analysis technique. We consider solid detection a critical step in the whole experimental procedure, which deserves special attention; the next three subsections are dedicated to it.

3.3.1. Automated Image Analysis. Since we are fully aware that human observation is also fallible, we have assessed its quality by comparing the values of $h_\text{d}$ determined by different human operators for the same set of images. We have observed a deviation among the different operators of a few percent, i.e., of about 3%, and thus we have decided to consider the human observation as a trustworthy reference train to validate the image analysis software.

The training uses two-thirds of the droplets analyzed by the human operator, $n_\text{train}$; the remaining droplets (one-third, i.e., $n_\text{test}$ with $n_\text{a} = n_\text{train} + n_\text{test}$) are used as a validation set, thus allowing for the evaluation of the so-called confusion matrix


given in the discussion thereof).

3.3.2. Empirical Evaluation of the Statistics of the Image Analysis Algorithm. The empirical approach is based on the following simple assumption: the number of droplets containing crystals that are detected by the image analysis algorithm, $h_\mu$, equals the number of droplets containing crystals, $h_\text{p}$, plus an error, $h_\epsilon$:

$$h_\mu = h_\text{p} + h_\epsilon$$  \hspace{1cm} (15)

All three quantities are uncorrelated random variables. Accordingly, the following relationships hold:

$$E[h_\mu] = E[h_\text{p}] + E[h_\epsilon]$$  \hspace{1cm} (16)

$$V[h_\mu] = V[h_\text{p}] + V[h_\epsilon]$$  \hspace{1cm} (17)

The three equations above constitute the model of the physical process comprising both physical experiment and automated image analysis. In order to use them, a model of the image analysis itself is needed, through which one estimates the statistical properties of the error $h_\epsilon$, appearing in eqs 16 and 17. This is obtained by using the results of the validation test of the image analysis algorithm (using $n_\text{test}$ droplets and corresponding images) and its $N_\text{nt}$ random combinations obtained through the bootstrap procedure. If $\Delta h_\text{test}$ is the difference between estimated (by the algorithm) and observed (by the human operator) droplets containing crystals out of the $n_\text{test}$ droplets, then $\Delta h_\text{test}/n_\text{test} = y_\text{d} - y_\text{e}$ (see section 3.3.1). Through the expected value and variance of the right-hand side of this identity, one can estimate the statistical properties of the error $h_\epsilon$ appearing in eqs 16 and 17 as
\[ E[h_i] = nE[y_i - y_e] = n\mu_e \]
\[ V[h_i] = n^2 V[y_i - y_e] = n^2\sigma_e^2 \]

where \( \mu_e \) and \( \sigma_e^2 \) are expected value and variance, respectively, of the difference \( y_i - y_e \), and are known from the bootstrap procedure. One can then rewrite eqs 16 and 17 as
\[
E[h_i] = n(\mu_p + \mu_e) \\
V[h_i] = n\mu_p (1 - \mu_p) + n(n - 1)\sigma_p^2 + n^2\sigma_e^2
\]

One can now estimate the expected value of \( h_i \), called \( \hat{h}_i \). This is related to the estimated value of \( \mu_p \), called \( \hat{p} \), because \( \hat{p} = \hat{h}_i/n \). In fact the expected value of \( h_i \) can be estimated through its measured value, called \( \hat{h}_i \) and obtained via the automated image analysis algorithm. Equations 20 and 21 yield
\[
\hat{h}_i = \hat{h}_i - E[h_i] = \hat{h}_i - n\mu_e \\
\hat{p} = \hat{h}_i/n
\]

where \( \hat{h}_i \) follows immediately using eqs 12 and 22 and using the property that the variance of the expected value of a random variable is the variance of that random variable divided by the number of data points, on which the expected value is calculated, i.e., \( V[E(h_i)] = V(h_i)/N_{bs} \) in this case:
\[
V[\hat{h}_i] = V[\hat{h}_i] + V[E(h_i)] \\
= n\hat{p}(1 - \hat{p}) + n(n - 1)\sigma_p^2 + \sigma_e^2\left(1 + \frac{1}{N_{bs}}\right)
\]

where \( \mu_e \) in eq 21 has been substituted by its estimate \( \hat{p} \) and \( \sigma_e^2 \) is calculated for a \( \mu \), whose mean is given by \( \hat{p} \). For the variance of \( \hat{h}_i \), one obtains from eq 24
\[
V[\hat{p}] = \frac{\hat{p}(1 - \hat{p})}{n} + \frac{n - 1}{n}\sigma_p^2 + \sigma_e^2\left(1 + \frac{1}{N_{bs}}\right)
\]

Equation 23 is plotted in Figure 4A for a value of \( \mu_e = 0.018 \), which is calculated as the average of the mean errors over the 42 experiments that have been carried out and evaluated. The average variance among all 42 experiments is \( \sigma_e^2 = 3.5 \times 10^{-3} \), which has to be compared with the values plotted in Figure 3 (solid lines). Since \( N_{bs} \) is quite large, the variance of the estimate of the success probability \( \hat{p} \) can be approximated very well as \( V[\hat{p}] \approx \hat{p}(1 - \hat{p})/n + \sigma_e^2 \).

In Figure 4A, the points corresponding to the 42 experiments are also plotted; these are close to the straight line, but not necessarily on it, because each experiment has its own value of \( \mu_e \). It is worth noting that eq 23 does not map the interval \((0, 1)\) of \( \hat{p} \) (note that \( 0 \leq \hat{p} = \hat{h}_i/n \leq 1 \) by definition) onto the interval \((0, 1)\) of \( \hat{p} \), because of the systematic detection error \( \mu_e \). Error bars in both \( \hat{p} \) and \( \hat{h}_i \) are also shown for some exemplary points, so as to highlight that the error is different in different experiments because of different values of \( \hat{p} \) and of the number of droplets \( n \).

Additionally, the empirical approach described here includes the special case where the automated image analysis is 100% accurate. In this particular case, which is idealized and unrealistic but relevant from a theoretical point of view as we see in the following, \( \mu_e = \sigma_e^2 = 0 \); hence, the straight line in Figure 4A coincides with the diagonal on which all experimental points fall.

3.3.3. Analytical Evaluation of the Statistics of the Image Analysis Algorithm. The analytical approach tries to capture the specific features of the image analysis algorithm: namely, the fact that it does not detect crystals in all droplets containing at least one and that sometimes detects that there are crystals in droplets that have none. Such features are represented by two quantities: \( w_{tp} \), fraction of true positives, and \( w_{fp} \), fraction of false positives, respectively.

As a consequence of the definitions above, the number of crystals containing droplets detected is
\[
h_j = h_{tp} + (n - h_j)w_{fp}
\]

All quantities here are random variables. The corresponding relationship in terms of expected values is obviously
\[
E[h_j] = E[h_{tp}]E[w_{tp}] + (n - E[h_j])E[w_{fp}]
\]

The variance of \( h_j \) is calculated, by assuming that the random variables in the two multiplications on the r.h.s. of eq 27 are independent, as
\[
V[h_j] = E[h_{tp}]^2V[w_{tp}] + E[w_{tp}]^2V[h_j] + (n - E[h_j])^2 \\
V[w_{tp}] + E[w_{tp}]^2V[h_j] + V[h_j](V[w_{tp}] \\
+ V[w_{fp}])
\]

where \( E[h_j] \) and \( V[h_j] \) are given by eqs 11 and 12 obeying the physical process of nucleation in microdroplets of distributed volumes.

There is a dual representation of the affine relationship between \( h_j \) and \( h_j \), as eq 26, namely
\[
h_j = w_1h_j - w_2
\]

where \( w_1 = 1/(w_{tp} - w_{fp}) \) and \( w_2 = w_{tp}/(w_{tp} - w_{fp}) \) and are also random variables. This can also be written in terms of expected values:
\[
E[h_j] = E[w_1]E[h_j] - nE[w_2]
\]

Equations 26–30 constitute the model of the physical process comprising both physical experiment and automated image analysis. Their use is possible when a model of the image analysis algorithm is available. This allows estimating the statistical properties of the random variables \( w_{tp}, w_{fp}, w_1 \), and \( w_2 \) appearing in those equations. This is obtained by using the

---

**Figure 4.** Parity plot of \( \hat{p} \) and \( \hat{h}_i \). Markers correspond to values of the 42 experiments; these are not necessarily on top of the dashed lines. Each experiment has its own variance, represented by the x.p error bars plotted for selected experiments. (A) Empirical evaluation of the statistics of the image analysis algorithm. The straight line is defined by the eq 23 for a value of \( \mu_e = 0.018 \). (B) Analytical evaluation of the statistics of the image analysis algorithm. The straight line is defined by eq 33 for values of \( E[w_1] = 1.91 \) and \( E[w_2] = 0.41 \).
results of the validation test of the image analysis algorithm (using \( n_{\text{test}} \) droplets and corresponding images) and its \( N_{\text{bs}} \) random combinations obtained through the bootstrap procedure. The estimated values and variances of the four coefficients above, i.e., \( w_{1p}, w_{2p}, w_{1y}, \) and \( w_{2y} \) can be computed through the estimated values and variances of \( x_{1p}, x_{2p}, z_{1}, \) and \( z_{2} \) obtained from the bootstrap procedure, because the latter and the former have the same physical meaning, though the former refer to the whole set of \( n \) droplets, whereas the latter refer to the subset \( n_{\text{test}} \).

As in the case of the empirical approach, one estimates the expected value of \( h_{p} \) using the actual value of \( h_{d} \) obtained by the image analysis algorithm. Then, one uses eq 31 to estimate the value of \( h_{p} \)

\[
\hat{h}_{p} = E[w_{1}]h_{d} - nE[w_{2}]
\]  

(31)

from which one can calculate the variance of \( \hat{h}_{p} \) as follows (after noting that \( V[E[w_{i}]] = \frac{V[w_{i}]}{N_{\text{bs}}}, i = 1, 2 \)):

\[
V[\hat{h}_{p}] = \hat{h}_{d}^{2} V[w_{1}] + E[w_{1}]^{2} V[\hat{h}_{d}] + \frac{V[w_{1}]}{N_{\text{bs}}} + \frac{n \varphi V[w_{1}]}{n^{2} N_{\text{bs}}}
\]  

(32)

The quantity of physical interest in this analysis is the mean of the distribution of success probabilities, \( \mu_{h_{p}} \), whose estimate is defined as \( \hat{p} = \hat{h}_{p}/n \). One can calculate it by dividing both terms of eq 31 by \( n_{i} \) thus obtaining the following expression, where—as in the empirical approach—the estimated quantity \( \hat{h}_{d} = h_{d}/n \) is used:

\[
\hat{p} = E[w_{1}]\hat{h}_{d} - E[w_{2}]
\]  

(33)

Finally, one calculates the variance of the estimate \( \hat{p} \) as

\[
V[\hat{p}] = E[w_{1}]^{2} V[\hat{h}_{d}] + \frac{\hat{h}_{d}^{2} V[w_{1}]}{n} + \frac{V[w_{1}]}{N_{\text{bs}}} + \frac{V[w_{1}]}{n^{2} N_{\text{bs}}}
\]  

(34)

The variance of \( \hat{h}_{d} \) is evaluated using eq 30 with \( E[\hat{h}_{d}] = n\hat{p} \) and \( V[\hat{h}_{d}] \) from eq 12 with \( \mu_{h} \) substituted by \( \hat{p} \) and \( \sigma_{E} \) calculated for a distribution \( \pi(p) \), whose mean is \( \hat{p} \).

The straight line defined by eq 33 is plotted in Figure 4B for values of \( E[w_{1}] \) and \( E[w_{2}] \) of 1.91 and 0.41, respectively, which are representative of the average of these quantities among the 42 experiments carried out and evaluated. The points corresponding to the 42 experiments are also plotted; they are close to the straight line, but not on it, because the two expected values above are different in each experiment. It is noteworthy that according to eq 34 the interval \((0, 1)\) of \( \hat{p} \) maps onto a much broader interval of \( \hat{s} \), as clearly illustrated in Figure 4B. This is due to the relatively large values of the two expected values above. On the one hand, even when there are no droplets containing particles, i.e., when \( \hat{p} = 0 \), due to the detection error in terms of false positives the expected value of \( \hat{s} \) is \( E[w_{2}]/E[w_{1}] = 0.22 \). On the other hand, when all droplets contain crystals, i.e., \( \hat{p} = 1 \), due to the detection error in terms of true positives, the expected value of \( \hat{s} \) is \((1 + E[w_{2}])/E[w_{1}] = 1.22 \).

Error bars in both \( \hat{s} \) and \( \hat{p} \) are also shown for some exemplary points. It is remarkable that the error bars are very large for experimental points where \( \hat{s} \) is close to its lower and upper bound and \( \hat{p} \) attains unphysical values, i.e., below 0 and above 1. This is a consequence of the functional form of eq 34 and of the fact that the statistics of the random variables \( w_{1} \) and \( w_{2} \) estimated through the bootstrap procedure show that these variables exhibit larger uncertainty in the limit cases where either only a few droplets contain crystals (\( \hat{p} \) close to 0) or most of the droplets contain crystals (\( \hat{p} \) close to 1).

3.3.4. Comparison of the Two Approaches. The empirical and analytical approaches to the evaluation of the statistics of the automated image analysis algorithm can be compared by plotting for the 42 experiments carried out and evaluating the values of the estimated success probability \( \hat{p} \) and the values of the variance associated with such estimate. This is shown in Figure 5, where the blue and red lines and symbols refer to the analytical and empirical evaluations, respectively. The symbols represent values from the experiments. It is worth noting that for each experiment different values not only of the variance but also of \( \hat{p} \) are obtained when the two different methods are applied and that when the values of \( \hat{p} \) are negative or larger than 1, they have been plotted as 0 and 1, respectively. The variance obtained applying the analytical method is always larger than that obtained with the empirical method. This is confirmed by the two curves that are calculated using eq 25 (red curve, for the empirical approach) and eq 34 (blue curve, for the analytical approach). The curves are calculated using average values of the expected values and variances of both image analysis features; these are found in Table S1 in the Supporting Information.

Figure 5. Variances of \( \hat{p} \). Blue and red solid curves and markers (connected by dashed lines for visual clarity) refer to the analytical and empirical evaluations of the image analysis statistics. Curves are calculated with average values of the expected values and variances of both image analysis features; these are found in Table S1 in the Supporting Information.

4. EVALUATION OF NUCLEATION EXPERIMENTS IN MICRODROPLETS

In this section we report the experimental results and analyze first the internal consistency of each experiment, as explained in section 4.1. Then, the experiments that are indeed internally
consistent are used to estimate nucleation rates and growth times, as discussed in detail in section 4.2.

4.1. Consistency and Reproducibility. The nucleation experiments in microdroplets presented and analyzed in this work involve thousands of droplets and have a duration of thousands of seconds. In an analysis of the experimental evidence it is necessary to assume that the nucleation probability (i.e., the success probability) is the same in all droplets (apart for the volume effects discussed in section 3.2), which implies that experimental conditions when droplets are formed are indeed the same for the whole duration of a single experiment. It is a legitimate question whether all droplets in a single experiment can indeed be considered similar in terms of nucleation probability or whether the droplets generated and reaching the end of the microfluidic device first are indeed similar to the droplets generated and observed last. On answering these questions hinges the possibility of judging whether two repeated experiments, i.e., carried out at the same nominal conditions of supersaturation, temperature, and droplet volume, which exhibit nevertheless different values of $\hat{s}$, could be considered comparable or not from a statistical viewpoint: i.e., could be considered belonging to the same stochastic process.

In order to answer these questions, we have devised a protocol to analyze the outcome of each individual experiment, as illustrated in Figure 6, where the horizontal coordinate is the number of droplets considered (on a logarithmic scale) and the vertical coordinate is the value of $\hat{s}$ obtained for that (sub)set of droplets (each of the eight subfigures corresponds to a different experiment).

The first step involves slicing the set of all droplets obtained in the specific experiment into subsets of smaller and smaller sizes and determining the value of $\hat{s}$ for each of the subsets. In the case of the experiment illustrated in Figure 6A (upper left plot), the whole group consisted of 3600 droplets and exhibited a value $\hat{s} = 0.33$ (rightmost symbol in the figure).

The original set was divided into two subsets of 1800 droplets each, i.e., the first 1800 droplets and the last 1800 droplets, each yielding its own $\hat{s}$ value: they are represented by the two symbols vertically aligned at $n_{\text{slice}} = 1800$ in the figure. Then the original is further divided—going from right to left—into three subsets of 1200 droplets each, and so on and so forth, all the way to 36 subsets containing 100 droplets each. Each subset represents an independent Bernoulli process that should be characterized by the same underlying success probability if the whole experiment consisting of 3600 droplets were indeed internally consistent. As the original set of 3600 droplets is sliced more and more, the estimate of the success probability is based on a smaller and smaller number of droplets and the associated variance (and the size of the confidence interval) becomes larger and larger. As a consequence, the 36 symbols representing the subsets with $n_{\text{slice}} = 100$ occupy a vertical segment, which is much longer than the confidence interval associated with the $\hat{s}$ value at $n = 3600$.

We can now define a criterion to establish if an experiment is internally consistent (the second step of our protocol). An experiment is internally consistent if the segments obtained as described above at all different values of $n_{\text{slice}}$ from 100 to the total number of droplets in an experiment, are contained in the shaded conelike region, also depicted in Figure 6A. Such a region is calculated by applying the following confidence intervals, $\pm \epsilon$, to the $\hat{s}$ value estimated for the entire droplet population of that experiment:

$$
\epsilon = \pm \frac{1}{2n_{\text{slice}}} \ln \frac{2}{\alpha}
$$

where $\alpha = 1 - C$ and $C$ is the confidence level (we have used $C = 0.95$ and $\alpha = 0.05$ in all cases). Note that the confidence interval depends only on the number of droplets in the subset of droplets and not on the estimated success probability (contrary to that obtained for a Bernoulli process). The shaded conelike region in Figure 6A accounts for the statistical features of any stochastic process, particularly of a Bernoulli process or of a sum of a number of independent nonidentical Bernoulli trials; it shows that there is an increase of uncertainty associated with a decrease in the number of trials and it quantifies it.

The confidence intervals of eq 35 are calculated using Hoeffding inequality,18,19 which gives a lower bound to the probability that the estimated value of a stochastic variable, $\hat{s}$ in our case, differs in absolute value from its expected value by a quantity smaller than $\epsilon$:

$$
\Pr(|\Delta \hat{s} < \epsilon|) \geq 1 - 2e^{-2\epsilon^2}
$$

Equation 35 is obtained from eq 36 by setting $n = n_{\text{slice}}$ and $\Pr(|\Delta \hat{s} < \epsilon|) = C$ and solving for $\epsilon$.

Therefore, the fact that the $\hat{s}$ estimates spread vertically as $n_{\text{slice}}$ decreases, e.g., from 3600 to 100 in Figure 6A, while remaining within the conelike region defined by the Hoeffding

![Figure 6](https://example.com/figure6.png)

**Figure 6.** Evaluation of internal consistency and reproducibility of single experiments. The rightmost symbol in each of the subplots is the $\hat{s}$ value for the original population of droplets. This is further sliced into smaller subsets—going from right to left—down to 100 droplets. The shaded conelike regions are calculated by applying the confidence intervals of eq 18 to the estimated values of $\hat{s}$ for the entire population. Each set of plots in the three subfigures shows repetitions of experiments carried out under the same operating conditions: (A) $S = 2.28$ and $\tau = 972.9$ s; (B) $S = 2.08$ and $\tau = 1557$ s; (C) $S = 2.37$ and $\tau = 681.1$ s.
inequality, indicates that all Bernoulli trials are (with 0.95 confidence level) statistically representative of the same underlying physical phenomena, as requested by internal consistency. In this study, we have carried out 45 experiments under different operating conditions, and all of them have fulfilled the criterion for internal consistency.

Let us now consider repetitions of the same experiment: e.g., the three experiments in Figure 6A (at $S = 2.28$ and $r = 972.9$ s), the two in Figure 6B (at $S = 2.08$ and $r = 1557$ s), and the three in Figure 6C (at $S = 2.37$ and $r = 681.1$ s). We use again the estimates obtained using all droplets in each experiment and the confidence intervals defined by eq 35 to assess whether the different experiments carried out under the same experimental conditions can be considered as representing the same physical phenomena, and hence, whether they can be used to estimate comparable values of the nucleation rate.

In the three cases illustrated in Figure 6 the different repetitions yield slightly different values of the estimates, but the Hoefling concenline regions for the different repetitions exhibit a major overlap, interpreted as the fulfillment of a reproducibility criterion. When two repeated experiments at the same nominal conditions do not overlap, we consider one of the two an outlier, which has to be excluded from further analysis. Out of 45 experiments, three of them (not shown here for brevity) have not fulfilled such a criterion and have been excluded, thus leaving 42 experiments (i.e., about 93% of the total) that have been analyzed further as discussed in the following. We were unfortunately not able to determine the root cause of the anomalous behavior of the three outliers.

### 4.2. Estimation of the Parameters in the Constitutive Equations Relevant to Nucleation in Microdroplets

In order to accomplish the overall objective of this study, namely the evaluation of nucleation rates and related quantities from the microfluidic experiments analyzed as discussed above, one needs several ingredients: (1) the set of estimated values of probabilities, $\hat{p}$, and associated variances, $V[\hat{p}]$, obtained from experiments through the procedures discussed above, (2) the physical model of the process to be characterized, which in this case links the observed success probability to the underlying nucleation rate, (3) the constitutive equations, i.e., giving the nucleation rate, $J$, and the growth time, $\tau_g$, in terms of supersaturation, temperature, and droplet volume, (4) the optimization procedure to estimate the parameters in the constitutive equations and the associated methodology to evaluate the variance and the confidence intervals of the estimated parameters. In this section, three sets of data will be considered, namely those obtained from the 42 experiments when the 3 approaches to the statistical evaluation of the image analysis algorithm are used: namely, those called empirical (section 3.3.2) and analytical (section 3.3.3) and the approach where we assume that the automated image analysis is error-free (brieﬂy discussed in section 3.3.2, and denominated for brevity “error-free” in the following). To each of these, two parameter regression methods will be applied, consisting of one or two steps, in both cases based on the weighted nonlinear least-squares method. Thus, six different sets of model parameters (and related uncertainties) will be obtained and comparatively assessed.

#### 4.2.1. Data

The $m$ experiments in this work, with $m = 42$, have been carried out at four different supersaturation levels, $S_j$, $j = 1, \ldots, 4$ (which is defined as the ratio of actual concentration to solubility at the experimental temperature, assuming that the activity coefficients at the two concentration levels are equal); at each $S_j$ level, $m_j$ experiments have been carried out (with $m_1 + m_2 + m_3 + m_4 = m$). The values of $\hat{p}$ obtained by applying the three different approaches to image analysis are presented in Table S2 in the Supporting Information and plotted in Figure 7 as a function of the residence time; the subfigures in the same column are obtained using the same approach, whereas the subfigures in each row correspond to the same supersaturation level, $S_j$. The error bar shown for every experimental value of $\hat{p}$ is calculated as the standard deviation from the corresponding variance $V[\hat{p}]$: i.e., its square root. Whenever automated image analysis delivers a value of $\hat{p}$ below 0 or above 1, $\hat{p}$ is assigned the value 0 or 1, respectively. The three approaches to assess the quality of the automated image analysis algorithm yield different, though similar, values of $\hat{p}$ but very different values of the standard deviation: i.e., very small (almost not visible) for the error-free approach, small for the empirical approach, and larger in the case of the analytical approach.

![Figure 7](https://example.com/figure7.png)
Table 2. Nucleation Rates, $\mathring{J}$, and Growth Times, $\mathring{\tau}$, Estimated with Eq 40 via the 2-Step Method with $\mathring{p}$ and $V[\mathring{p}]$ Stemming from the Error-Free, Empirical and the Analytical Treatments of the Image Analysis Statistics and Parameters of the Nucleation Rate, $a$ and $b$, and Parameters of the Growth Time, $\gamma$ and $\delta$, Estimated, either Using $J$ and $\mathring{\tau}$ Values, Obtained Using Nominal Supersaturations, $S$, and Mean Volumes, via the 2-Step Method, or Obtained Applying Eq 45 Directly for All Supersaturations, $S$, via the 1-Step Regression$^a$.

| $S$ | $J$ ($10^{-7}$ m$^3$ s$^{-1}$) | $\mathring{\tau}$ (s) | $a$ | $b$ | $\gamma$ | $\delta$ |
|-----|------------------|-----------------|-----|-----|---------|---------|
|     |      |              |     |     |         |         |
| 2-step |     |              |     |     |         |         |
| 2.08 | 0.92 ± 0.18 | 11539 ± 37.3 | 0.36 ± 1.18 | 1.86 ± 0.80 | 1.40 ± 0.31 | 2.36 ± 1.19 |
| 2.18 | 1.59 ± 0.41 | 9887 ± 58.1 |          |         |         |         |
| 2.28 | 1.94 ± 0.42 | 8364 ± 36.4 |          |         |         |         |
| 2.37 | 3.44 ± 1.49 | 6150 ± 40.5 |          |         |         |         |

1-step$^b$

| $S$ | $J$ ($10^{-7}$ m$^3$ s$^{-1}$) | $\mathring{\tau}$ (s) | $a$ | $b$ | $\gamma$ | $\delta$ |
|-----|------------------|-----------------|-----|-----|---------|---------|
|     |      |              |     |     |         |         |
| 2-step |     |              |     |     |         |         |
| 2.08 | 0.82 ± 0.26 | 11057 ± 73.3 | 0.48 ± 1.23 | 1.96 ± 0.85 | 1.36 ± 0.35 | 2.16 ± 1.389 |
| 2.18 | 1.62 ± 0.46 | 9738 ± 66.7 |          |         |         |         |
| 2.28 | 1.98 ± 0.34 | 8278 ± 30.3 |          |         |         |         |
| 2.37 | 2.92 ± 1.53 | 581.3 ± 67.1 |          |         |         |         |

1-step$^b$

| $S$ | $J$ ($10^{-7}$ m$^3$ s$^{-1}$) | $\mathring{\tau}$ (s) | $a$ | $b$ | $\gamma$ | $\delta$ |
|-----|------------------|-----------------|-----|-----|---------|---------|
|     |      |              |     |     |         |         |
| 2-step |     |              |     |     |         |         |
| 2.08 | 0.77 ± 0.39 | 11433 ± 104.6 | 2.11 ± 2.76 | 2.87 ± 1.94 | 1.60 ± 0.42 | 2.79 ± 1.47 |
| 2.18 | 2.14 ± 0.85 | 1057.3 ± 58.1 |          |         |         |         |
| 2.28 | 2.44 ± 0.69 | 858.3 ± 38.1 |          |         |         |         |
| 2.37 | 4.99 ± 2.19 | 623.5 ± 35.1 |          |         |         |         |

1-step$^b$

| $S$ | $J$ ($10^{-7}$ m$^3$ s$^{-1}$) | $\mathring{\tau}$ (s) | $a$ | $b$ | $\gamma$ | $\delta$ |
|-----|------------------|-----------------|-----|-----|---------|---------|
|     |      |              |     |     |         |         |
| 2-step |     |              |     |     |         |         |
| 2.08 | 0.76 ± 4.80 | 144 ± 2.99 | 1.28 ± 0.52 | 2.67 ± 2.05 |

$^a$The errors reported are the 95% confidence intervals. $^b$ $a$, $b$, $\gamma$, and $\delta$ values are for $S$, for the 1-step method.

4.2.2. Physical Model. The mean of the success probability $\mu_k$ is related to the nucleation rate $J$ through the relationship of eq 8, that can be recast as

$$
\mu_k = 1 - \left(1 + \frac{\mu J(\tau - \mathring{\tau})}{\alpha}\right)^{-\alpha}
$$

where $\mu$, and $\alpha$ are the mean volume and squared reciprocal of the coefficient of variation of the distribution of droplet volumes, respectively. In this equation $J$ and $\mathring{\tau}$ depend on the supersaturation, $S$; the mean probability $\mu_k$ depends also on the residence time $\tau$ and on the droplet volume $\mu_c$.

4.2.3. Constitutive Equations. The nucleation rate is expressed through classical nucleation theory, i.e., using the following standard two-parameter relationship:

$$
J = A_0 S \exp\left(\frac{a - \beta}{\ln^2 S}\right)
$$

where $a$ and $\beta$ are the dimensionless parameters to be estimated, $A_0 = 1 \times 10^6$ m$^3$ s$^{-1}$ is a constant reference value of the pre-exponential coefficient $A$, and $a = \ln(A/A_0)$. It is worth noting that by choosing this nucleation rate expression we are assuming that the phenomenon we are observing is primary nucleation. Ascertainment whether this is homogeneous or heterogeneous, e.g., at the interface between continuous and dispersed phase, is beyond the scope of this work; eq 38 can describe both.

For the growth time the following two-parameter semi-empirical relationship is used:

$$
\mathring{\tau} = \frac{C_0 \gamma}{(S - 1)^{\delta}}
$$

where $\gamma$ and $\delta$ are positive dimensionless parameters to be estimated; $C_0 = 1 \times 10^3$ s is a constant reference value of the numerator. The last equation expresses the growth time as the ratio between a characteristic length and an empirical growth rate, proportional to the driving force for growth, namely $S - 1$, to the power of $\delta$; the positive model parameters guarantee that $\mathring{\tau}$ decreases with increasing $S$.

4.2.4. Variances and Confidence Intervals of the Estimated Parameters. Approximate values of the variance and of the confidence intervals of the parameters estimated through the two methods presented below, called here for the sake of simplicity $\beta_0$, can be determined as follows. One calculates the sensitivity matrix: i.e., the matrix of the derivatives of the model predictions with respect to the model parameters, which are based on a linearization of the model in the vicinity of the estimated parameter values. Then one uses the sensitivity matrix to estimate the covariance matrix of the parameter estimates, by exploiting the weighted sum of the residuals. The variance of the $k$th model parameter, $V[\beta_k]$ is given by the $k$th diagonal element of such covariance matrix, whereas the standard deviation is obviously its square root. The confidence interval of the parameter $\beta_k$ with a confidence of, e.g., 95% is then given by its standard deviation multiplied by the value of the $t$ distribution for a 95% confidence and a number of degrees of freedom equal to the number of data points minus the number of parameters.

4.2.5. Parameter Estimation via a Two-Step Method. The first method consists of two optimization steps. First, the $\beta_i$,...
experiments carried out at each of the four supersaturation levels, \( S_p \), are used to estimate the corresponding nucleation rate \( J \) and growth time \( \tau \), and the associated variances. Then the four values of these quantities thus obtained are exploited to estimate the two plus two parameters in the corresponding constitutive equations. In all cases we use the \texttt{fminsearch} function as implemented in the optimization toolbox in Matlab.

Therefore, in the first step the optimization problem involving the minimization of the following weighted sum of the squared residuals is solved:

\[
\text{minimize}_{J,\tau} \sum_{i=1}^{m} \frac{(\hat{p}_{j,i} - \tilde{p}_{j,i}(J, \tau))}{V[\hat{p}_{j,i}]}^2
\]

where \( \hat{p}_{j,i} \) is the value of the mean success probability in the \( i \)th experiment of the series at \( S_j \) (with \( V[\hat{p}_{j,i}] \) being its variance) and \( \tilde{p}_{j,i} \) is the corresponding predicted value, defined as

\[
\tilde{p}_{j,i} = 1 - \left( 1 + \frac{\mu_{\alpha,i} J_{v,i}(\tau_j - \tau)_{g,i}}{\alpha_{j,i}} \right)^{-\gamma_{j,i}}
\]

It is worth noting that the values of \( \mu_{\alpha}, \alpha, \) and \( \tau \) are known, though specific to the \( i \)th experiment of the series at \( S_p \) as indicated by the subscripts.

This optimization is performed four times for the four supersaturation levels, thus yielding the estimated values of nucleation rate, \( J \) and \( \tau \), and the associated variances \( V[J] \) and \( V[\tau] \), calculated through the procedure summarized in section 4.2.4. Such procedure is repeated for each approach in evaluating the experimental data: i.e., the empirical, analytical, and error-free evaluation of the automated image analysis algorithm. The results are reported in detail in Table 2 and plotted in Figure 8 (where the columns refer to the three different approaches to assess image analysis, whereas the first and second rows are for nucleation rate and growth time, respectively).

\[
\sum_{i=1}^{m} \frac{(\hat{p}_{j,i} - \tilde{p}_{j,i}(a, b, \gamma, \delta))}{V[\hat{p}_{j,i}]}^2
\]

where the values of \( \tilde{J} \) and \( \tilde{\tau} \) (together with their variances) are the outcome of step 1 and the predicted values \( \tilde{J} \) and \( \tilde{\tau} \) are calculated using eqs 38 and 39, with \( S = S_p \).

These two optimizations yield the estimated values of the model parameters \( \hat{a}, \hat{b}, \hat{\gamma}, \) and \( \hat{\delta} \), with the associated variances and confidence intervals (see Table 2 for all of the values obtained). The results are illustrated in two ways. On the one hand, the constitutive relationships eqs 38 and 39 are plotted in Figure 8 as solid lines, using the estimated values of the model parameters. On the other hand, such values are used to draw the solid lines in Figure 7; they differ from the corresponding dashed lines, because of the deviation in predicting the experimental values of \( J \) and \( \tau \), the fitted values of the constitutive equations and the estimated model parameters \( \hat{a}, \hat{b}, \hat{\gamma}, \) and \( \hat{\delta} \).

4.2.6. Parameter Estimation via a One-Step Method. The second parameter estimation method consists of a single step where the four model parameters in the constitutive equations are estimated altogether, by solving the following optimization problem:

\[
\text{minimize}_{a,b,\gamma,\delta} \sum_{i=1}^{m} \frac{(\hat{p}_{j,i} - \tilde{p}(a, b, \gamma, \delta))}{V[\hat{p}_{j,i}]}^2
\]

where \( \tilde{p}_{j,i} \) is the value of the mean success probability in the \( i \)th experiment (with \( V[\tilde{p}_{j,i}] \) being its variance), and \( \tilde{p}_{j} \) is the corresponding predicted value, defined as

\[
\tilde{p}_{j,i} = 1 - \left( 1 + \frac{\mu_{\alpha,i} J_{v,i}(\tau_j - \tau)_{g,i}}{\alpha_{j,i}} \right)^{-\gamma_{j,i}}
\]

where the known values \( \mu_{\alpha,i} \) and \( \alpha_{j,i} \) are those characterizing the droplet volume distribution, supersaturation and the residence time, respectively, in the \( i \)th experiment. These values are reported in Table S2 in the Supporting Information.

The solution of the optimization problem above (once for each of the three approaches to the characterization of the image analysis algorithm) yielded the estimated values of the
model parameters $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, and $\tilde{d}$, as well as their variances and confidence intervals (see Table 2). We have also compared the values of the objective function of eq 44, calculated using these values of the model parameters, with those of the same objective function calculated using the corresponding model parameters obtained through the 2-step method (which solve the minimization problem of eqs 40, 42, and 43). It is apparent that the former are always smaller than the latter, thus providing an indication that the four parameters obtained with the 1-step method correspond indeed to a global optimum. The outcome of the parameter estimation carried out using the 1-step method is illustrated in two figures. First in Figure 8 the constitutive relationships eqs 38 and 39 are plotted vs supersaturation as dashed-dotted lines, using the estimated values of the model parameters. Then, in Figure 7 the same values are used to draw as dashed-dotted lines the exponential distributions of nucleation times for the four experimental supersaturation levels.

4.2.7. Comparative Assessment. Let us first analyze more in depth the results obtained by applying the 2-step method. On the one hand the uncertainties on the experimental $p$ values shown in Figure 7 increase significantly on going from the error-free approach to the empirical and to the analytical approach. On the other hand, the uncertainties on the estimated nucleation rates are similar for the error-free and the empirical approaches, whereas they are significantly larger in the case of the analytical approach. Such uncertainties are primarily determined by the limited number of residence time values explored at each supersaturation level and only secondarily by the errors due to the automated image analysis method. Nevertheless, it is worth noting that even at the highest supersaturation level, where the estimated nucleation rates are quite different following the three different approaches and the error bars on the estimated nucleation rates are the largest (see Figure 8, top row), the confidence intervals calculated using the three different approaches overlap at least partially. In contrast, the values of growth time are remarkably similar when the different approaches are applied, with comparatively much smaller uncertainties than for nucleation rates.

As far as the values of the estimated model parameters reported in Table 2 are concerned, it is worth noting that neither $c$ nor $d$ changes sign in their entire 95% confidence intervals (whatever the approach applied to evaluate the automated image analysis is): i.e., they fulfill the physical constraints associated with the chosen functional form of $r_g$ given by eq 39. The nucleation rate parameter $b$ does not change sign either within the 95% confidence interval, which is also needed to fulfill the requirement that the nucleation rate increases for increasing supersaturation. In contrast, the nucleation rate parameter $a$ changes sign when it is varied within the whole 95% confidence interval, which is physically possible, since this does not affect the sign of the pre-exponential rate constant in the nucleation rate expression of eq 38. It can be readily observed that the estimated value $\tilde{a}$ is larger when it is estimated applying the analytical approach, which is compensated by a larger value of $\tilde{b}$.

Finally, let us consider the absolute values of the growth time and of the nucleation rate obtained using the three different approaches to evaluating the automated image algorithm (see Figure 8). On the one hand, the growth time values at the individual supersaturation levels, as well as the overall growth time correlation, are remarkably similar in the three cases. On the other hand, the individual nucleation rates are similar when they are obtained from the error-free and the empirical approach but rather different from the analytical approach (with differences between 25% and 50%). The nucleation rate correlation describes well the estimated values in all three cases, as observed in the top row of Figure 8. The differences among the three cases underline the importance of image analysis in the quantitative determination of nucleation rates using nucleation experiments in microdroplets. Image analysis has to be as accurate as possible, and its associated uncertainties have to be evaluated carefully. However, it is remarkable that even with very different assumptions about the quality of the automated image analysis algorithm the estimated values of the nucleation rate and of its correlation differ by less than a factor of 2 within the supersaturation interval explored.

Let us now consider the 1-step method, which is appealing at least from a theoretical viewpoint. With reference to both Figures 7 and 8, it is rather obvious that the curves obtained using the parameters estimated via the 1-step method are not able to describe the experimental data as accurately as those drawn based on the 2-step method. This is certainly due to the fact that in the 1-step case only 4 parameters are used to describe all 42 experiments, whereas in the 2-step case the data at each supersaturation level are fitted using 2 parameters, namely $J$ and $r_g$, and then the $J$ values and the $r_g$ values obtained in this way are regressed independently, each of the two sets using two parameters. Such an observation also justifies the fact that the confidence intervals on $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, and $\tilde{d}$ obtained through the 1-step method are larger than those on the same parameters estimated via the 2-step method, in fact much larger in the case of the nucleation rate parameters $\tilde{a}$ and $\tilde{b}$. However, in the case of $\tilde{b}$ such large uncertainty leads to confidence intervals within which the parameter $b$ may change sign and become negative, which is physically inconsistent, as it would lead to the wrong dependence of $J$ on $S$: namely, $J$ would decrease with increasing $S$, which is obviously unphysical.

5. CONCLUDING REMARKS

In this work we report a study of primary nucleation of adipic acid from microdroplets consisting of a supersaturated aqueous solution. The study has been carried out using a fully automated microfluidic setup (including a crystal detection procedure based on an automated image analysis algorithm), with which we were able to continuously generate and analyze thousands of monodispersed microdroplets. A careful statistical analysis of the experimental results has allowed us to assess how uncertainties propagate from the stochastic occurrence (or not) of nucleation in an individual droplet, to statistically characterize the ensemble of hundreds or thousands of virtually identical droplets in an individual experiment, and to estimate the nucleation rate and the growth time of the related model parameters. Such a comprehensive study has allowed us to obtain a number of novel results, which not only apply to the specific system studied but also bear a much more general validity.

From an experimental perspective we have devised ways to generate microdroplets exhibiting a very narrow droplet size distribution and to measure it, and we have developed a new protocol to verify the internal consistency of an individual experiment: i.e., a protocol that allows us to identify and eliminate possible outliers (only about 7% of our experiments).
Also from an experimental viewpoint, we have developed an automated image analysis procedure, which exploits training from the operator first (based on about 10% of all the droplets belonging to an individual experiment), and it is utilized thereafter to analyze all the droplets of the same experiment, thus enabling the slicing procedure described in section 4.1 and used to assess internal consistency of that experiment.

From a theoretical point of view, we have been able to characterize the statistics of the estimated value of the success (or nucleation) probability in each experiment in terms of expected value and variance. We have done this by accounting explicitly for the nature of the stochastic process (which is the sum of independent nonidentical Bernoulli trials), for the distribution of microdroplet volumes, and for the error of the automated image analysis procedure.

Whereas the variance associated with the Bernoulli process can be reduced by increasing the number of droplets (the variance of the estimated success probability is inversely proportional to the number of droplets on which the estimate is based), that due to the volume distribution is not. The latter is effectively negligible for droplet volume distributions as narrow as that in our experiments (coefficient of variation of about 2%) but becomes dominant if the coefficient of variation of the volume distribution exhibits values of e.g. 15% (not uncommon in the microfluidic literature on microdroplet generation).

The accuracy of image analysis is critical, not only because it tends to dominate uncertainty in the estimation of nucleation rates and related model parameters but also because such accuracy is very difficult to assess. In this context, we have considered three alternative approaches on how to evaluate it: namely, an analytical approach based on a description of the underlying physics and statistics of the image analysis algorithm, an empirical approach, and an idealized approach that assumes that image analysis is error-free (which is used for reference and represents a best-case scenario). These different approaches treat the outcome of the image analysis procedure as a random variable, whose expected value and variance follow from those of the physical process (nucleation in microdroplets) and of the image analysis algorithm; the latter statistics are obtained by applying the bootstrap procedure described in section 3.3.1. The variance of the estimated success probability decreases significantly on going from the analytical to the empirical and to the error-free approach.

We have shown how estimated success (nucleation) probabilities obtained at the same supersaturation but for different residence times can be used to estimate the nucleation rate and the growth time associated with that supersaturation. It is striking that the uncertainty on such physical quantities is less sensitive on the assumed accuracy of the image analysis procedure, because it is affected a great deal by the amplitude of the range of residence times explored (which is by definition limited, as one can readily understand by inspecting Figure 7) and by the steepness of the exponential distribution of success probabilities as a function of nucleation time (given by eq 37 and illustrated in the same Figure 7), which is a physical feature of the system.

Finally, we have estimated the two plus two model parameters appearing in the constitutive equations for the nucleation rate and the growth time. To this aim, we have tested two procedures and shown that the one consisting of two steps, i.e., estimating nucleation rates and growth times at each supersaturation level first, and then estimating the model parameters by fitting the two constitutive equations independently to those values, outperforms that based on a single step. This is true despite the fact that the estimated values of the parameters depend on which of the image analysis evaluation approaches is applied (though the obtained values differ one from the other by no more than 50%) and that the associated confidence intervals are rather large (though physically consistent).

The key conclusion of our analysis is that microfluidics can indeed be used to estimate nucleation rate values and nucleation rate relationships because it enables performing a huge number of nucleation experiments in microdroplets under the same controlled conditions. However, the accuracy of these estimates very much depends not only on the number of droplets but also on the droplet volume distribution, on the image analysis algorithm, and on the number and range of experimental conditions (residence times and supersaturation levels) explored. We have developed and reported exact equations that allow calculations of the statistics of how uncertainty propagates by accounting for all the effects above. We hope that researchers in the field, who are motivated to exploit the valuable features of microfluidics to characterize nucleation, will apply such relationships and will carefully report the uncertainty of their measurements and of their estimates.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.cgd.9b00562.

Exemplary figures including the scale of droplets containing crystals of the images recorded at the observation zone, values for estimating the uncertainty added to nucleation probability due to processing images with the classification decision software, and all experimental conditions (i.e., supersaturations, residence times, mean droplet volumes, shape factors of the volume distribution, and total number of droplets) (PDF)

**AUTHOR INFORMATION**

Corresponding Author

*M.M.: tel, +41 44 632 24 56; fax, +41 44 632 11 41; e-mail, marco.mazzotti@ipe.mavt.ethz.ch.

ORCID®

Giovanni Maria Maggioni: 0000-0002-8830-3473

Marco Mazzotti: 0000-0002-4948-6705

Present Address

§School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, United States.

Notes

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The authors are thankful to Agnieszka Ladosz for fruitful discussions, to Bernhard Roth for conducting proof of concept experiments of the experimental setup, to Daniel Trottmann for the crystallization and quench zone construction, to Jovo Vidic for setting up the IR sensors and adjusting its signal reception, to Ashwin Rajagopalan and Stefan Bötschi for their help with the image analyzer, and to Pietro Binel and Tuvshee...
Otgobayar for classifying a large training set of droplets. Furthermore, the authors gratefully acknowledge the financial support of the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement number 2-73959-18).

**NOTATION**

- $\tilde{a}$: estimated parameter of the nucleation rate $J$ expression
- $A$: tube cross section ($m^2$)
- $A_0$: reference value of the pre-exponential coefficient of the nucleation rate $J$ expression ($m^{-3} s^{-1}$)
- $\tilde{b}$: estimated parameter of the nucleation rate $J$ expression
- $c^o$: solubility of the solute in solvent (g/g)
- $C$: confidence level
- $C_0$: reference value of numerator of the growth time $\tilde{\tau}_g$ semiempirical expression
- $E$: operator of expected value
- $f$: frequency of droplet formation and flow (Hz)
- $F$: cumulative exponential probability distribution
- $f_{FP}$: fraction of true positives
- $f_{FN}$: fraction of false negatives
- $h_p$: number of droplets containing crystals/number of success for distributed droplets
- $h_o$: number of droplets containing crystals observed by the human expert
- $h_d$: number of droplets containing crystals detected by the image analyzer
- $h_e$: error of the droplets containing crystals detected by the image analyzer
- $\tilde{h}_d$: estimated expected value of $h_d$
- $\tilde{h}_p$: estimated expected value of $h_p$
- $J$: nucleation rate ($m^{-3} s^{-1}$)
- $L_{\text{Tube}}$: length of the crystallization zone (m)
- $m$: number of experiments
- $n$: number of droplets
- $n_o$: number of droplets classified by the human expert
- $n_{\text{test}}$: number of droplets classified as used in a training set for the image analyzer
- $n_{\text{train}}$: number of droplets classified used for validation of the image analyzer
- $n_{\text{slice}}$: subdivision of the number of droplets for assessing internal consistency
- $N_{\text{bs}}$: number of bootstrap made per experiment
- $P$: success probability
- $\tilde{P}$: estimated value of the success probability
- $P_r$: operator of probability
- $q_d$: flow rate of the dispersed phase ($\mu$L/min)
- $q_c$: flow rate of the continuous phase ($\mu$L/min)
- $Q$: total flow rate ($\mu$L/min)
- $\tilde{s}$: estimated value of the measured success probability
- $S$: supersaturation
- $t_{\text{exp}}$: duration of the experiment (s)
- $T$: temperature (°C)
- $U$: superficial velocity (mm/s)
- $V$: volume of each droplet (nL)
- $\tau$: residence time (s)
- $V_r$: operator of variance
- $w_{FP}$: ratio of $n$ droplets containing crystals correctly predicted
- $w_{FN}$: ratio of $n$ droplets wrongly predicted to contain crystals
- $w_1$: rearranged relation of $x_{FP}$ and $x_{TP}$ using $n_{\text{test}}$ droplets
- $w_2$: rearranged relation of $x_{TP}$ and $x_{FP}$ using $n_{\text{test}}$ droplets
- $x_{TP}$: ratio of the fraction of $n_{\text{test}}$ droplets containing crystals correctly predicted
- $x_{FP}$: ratio of the fraction of $n_{\text{test}}$ droplets wrongly predicted to contain crystals
- $y_o$: fraction of droplets containing crystals, as determined by the human expert
- $y_d$: fraction of droplets containing crystals, as detected by the image analyzer
- $z_1$: rearranged relation of $x_{TP}$ and $x_{FP}$ using $n$ droplets
- $z_2$: rearranged relation of $x_{TP}$ and $x_{FP}$ using $n$ droplets

**GREEK LETTERS**

- $\alpha$: shape parameter of the gamma volume distribution
- $\beta$: rate parameter of the gamma volume distribution ($m^{-3}$)
- $\gamma$: estimated parameter of the growth time $\tilde{\tau}_g$ semiempirical expression
- $\Gamma$: gamma function
- $\delta$: estimated parameter of the growth time $\tilde{\tau}_g$ semiempirical expression
- $\epsilon$: confidence intervals calculated using the Hoeffding’s inequality
- $\mu_p$: mean of the success probability distribution
- $\mu_v$: mean volume of a population of droplets (nL)
- $\mu_{y_d}^x$: expected value of the difference $y_d - y_o$
- $\phi(p)$: probability density function of success probability distribution
- $\sigma$: standard deviation of droplet volumes
- $\sigma^2$: variance of the success probability distribution
- $\sigma_{y_d}^2$: variance of the difference $y_d - y_o$
- $\tau$: residence time (s)
- $\tau_{FP}^x$: growth time (s)
- $\psi$: coefficient of variation

**LATIN LETTERS**

- $\Delta P$: pressure drop (mbar)
- $\Delta t$: time interval between two successive droplets (s)

**SUBSCRIPTS AND SUPERSCRIPTS**

- $i$: index of number of experiments
- $j$: index of number of supersaturations employed
- $k$: running variable of number of droplets

**REFERENCES**

(1) Laval, P.; Crombez, A.; Salmon, J.-B. Microfluidic droplet method for nucleation kinetics measurements. *Langmuir* 2009, 25, 1836–1841.

(2) Teychene, S.; Biscans, B. Microfluidic device for the crystalization of organic molecules in organic solvents. *Cryst. Growth Des.* 2011, 11, 4810–4818.

(3) Toldy, A. I.; Badruddoza, A. Z. M.; Zheng, L.; Hatton, T. A.; Gunawan, R.; Rajagopalan, R.; Khan, S. A. Spherical crystalization of glycine from monodisperse microfluidic emulsions. *Cryst. Growth Des.* 2012, 12, 3977–3982.

(4) Dombrowski, R. D.; Litster, J. D.; Wagner, N. J.; Hey, Y. Crystallization of alpha-lactose monohydrate in a drop-based microfluidic crystallizer. *Chem. Eng. Sci.* 2007, 62, 4802–4810.

(5) Lu, J.; Litster, D.; Nagy, Z. K. Nucleation studies of active pharmaceutical ingredients in an air-segmented microfluidic drop-based crystallizer. *Cryst. Growth Des.* 2015, 15, 3645–3651.

(6) Rossi, D.; Gavrilidis, A.; Kuhn, S.; Candel, M. A.; Jones, A. G.; Price, C.; Mazzei, L. Adipic acid primary nucleation kinetics from

6173

DOI: 10.1021/acs.cgd.9b00562

Cryst. Growth Des. 2019, 19, 6159–6174
probability distributions in droplet-based systems under stagnant and flow conditions. Cryst. Growth Des. 2015, 15, 1784–1791.

(7) Ladosz, A.; Rigger, E.; von Rohr, P. R. Pressure drop of three-phase liquid-liquid–gas slug flow in round microchannels. Microfluid. Nanofluid. 2016, 20, 49.

(8) Wolffenbuttel, B. M. A.; Nijhuis, T. A.; Stankiewicz, A.; Moulijn, J. A. Novel method for non-intrusive measurement of velocity and slug length in two- and three-phase slug flow in capillaries. Meas. Sci. Technol. 2002, 13, 1540–1544.

(9) dos Santos, E. C.; Ladosz, A.; Maggioni, G. M.; von Rohr, P. R.; Mazzotti, M. Characterization of shapes and volumes of droplets generated in pdms t-junctions to study nucleation. Chem. Eng. Res. Des. 2018, 17, 2852–2863.

(10) Prileszky, T. A.; Oggunnaie, B. A.; Furst, E. M. Statistics of droplet sizes generated by a microfluidic device. AIChE J. 2016, 62, 2923–2928.

(11) Mullin, J. W. Crystallization; Oxford University Press: 2001.

(12) Rajagopalana, A. K.; Schneeeberger, J.; Salvatori, F.; Bötschi, S.; Ochsenbein, D. R.; Oswald, M. R.; Pollefeys, M.; Mazzotti, M. A comprehensive shape analysis pipeline for stereoscopic measurements of particulate populations in suspension. Powder Technol. 2017, 321, 479–493.

(13) Chen, K.; Goh, L.; He, G.; Kenis, P. J. A.; Zukoski, C. F.; Braatz, R. D. Identification of nucleation rates in droplet-based microfluidic systems. Chem. Eng. Sci. 2012, 77, 235–241.

(14) Brandel, C.; ter Horst, J. H. Measuring induction times and crystal nucleation rates. Faraday Discuss. 2015, 179, 199–214.

(15) Maggioni, G. M.; Mazzotti, M. Modelling the stochastic behaviour of primary nucleation. Faraday Discuss. 2015, 179, 359–382.

(16) Maggioni, G. M.; Bosetti, L.; dos Santos, E. C.; Mazzotti, M. Statistical analysis of series of detection time measurements for the estimation of nucleation rates. Cryst. Growth Des. 2017, 17, 5488–5498.

(17) Zhang, S.; Curien, C. G.; Veesler, S.; Candoni, N. Prediction of sizes and frequencies of nanoliter-sized droplets in cylindrical t-junction microfluidics. Chem. Eng. Sci. 2015, 138, 128–139.

(18) Wassermann, L. All of Nonparametric Statistics; Springer: 2006.

(19) Massart, P. The tight constant in the dvoletsky-kieler-wolfowitz inequality. Ann. Probab. 1990, 18, 1269–1283.