ABSTRACT: Nanoparticle size impacts properties vital to applications ranging from drug delivery to diagnostics and catalysis. As such, evaluating nanoparticle size dispersity is of fundamental importance. Conventional approaches, such as standard deviation, usually require the nanoparticle population to follow a known distribution and are ill-equipped to deal with highly poly- or heterodisperse populations. Herein, we propose the use of information entropy as an alternative and assumption-free method for describing nanoparticle size distributions. This measure works equally well for mono-, poly-, and heterodisperse populations and represents an unbiased route to evaluation and optimization of nanoparticle synthesis. We provide an intuitive tool for analysis with a user-friendly macro and provide guidelines for interpretation with respect to known standards.

Many characteristics of nanoparticles, including their biodistribution as well as catalytic, optical, and electric properties, depend on their spatial dimensions. Yet, accurate reporting on the size distribution remains a challenge, largely due to the absence of a reliable, widely applicable method of representing the dispersity of the population. While a histogram contains a full description of the particle size distribution, working with a scalar measure of dispersity is advantageous to determine correlations with properties, depend on their spatial dimensions. This need for a reliable descriptor of dispersity is particularly profound when optimizing syntheses to produce monodisperse nanoparticles; however, limited success was achieved in applying to a normal distribution and may provide an insufficient representation of the sample. Standard deviation is often reported against the mean size to produce the unit-less coefficient of variance (COV), which reflects the relative spread of the given distribution. Less commonly used metrics in nanoparticle science reporting include the percentile values, corresponding to the particle sizes that encompass 10, 50, and 90% of the total population (D10, D50, and D90 respectively). Using these percentile values, a representation of the dispersity can be calculated, (D90 − D10)/D50, referred to as the span. This is a more widely applicable measurement; it does not rely on a known distribution and can be employed even with polydisperse samples. Like the COV, the span is a relative measure, which relates the spread of the population to the mean or median value. These metrics focus on describing the breadth of the distribution. Dispersity represents the inhomogeneity of particle sizes observed in a population; while this typically correlates with the spread, it is not equivalent. For example, a population of nanoparticles with two distinct sizes will demonstrate a span and standard deviation dependent on the difference between the two sizes; the heterogeneity of sizes remains constant.

We propose the use of a modified version of the information entropy equation to accurately evaluate dispersity. Information entropy (H) was first proposed by Claude Shannon in 1948 to quantify the amount of information produced by a given process. It is calculated by

\[ H = -\sum_{i=1}^{n} p_i \ln(p_i) \]  

The probability of outcome i is denoted p_i, there are n possible outcomes. The entropy is often described as being analogous to the amount of information conveyed when the outcome of an event is observed. For example, if the result of a process is
absolutely certain, i.e., there is only one possible outcome \( (n = 1, p_1 = 1) \), no information is gained by observing the outcome as it is always the same and the entropy is found to be 0. As the uncertainty of the outcome increases, so does the entropy, and more information is revealed by the result.

This measure of entropy was soon adopted by a wide variety of areas of study outside of information theory including the study of species diversity, population genetics, molecular analysis, and finance.\(^5\text{\textendash}^8\) The pertinence of information entropy to such diverse fields lies in its ability to encapsulate not only the number of subcategories but also the relative quantities observed. It is not restricted to known distributions and can be applied to any data set. These properties make information entropy an ideal candidate to measure dispersity in nanoparticle populations.

**Methods**

Figure 1 illustrates an example of the entropy calculation. Here, 35 particles of 7 different possible sizes are sorted into appropriate bins (or intervals) of a histogram. The number of particles in each bin is divided by the total to determine the probability distribution.

![Figure 1. Nanoparticle entropy. (a) Schematic of nanoparticle population with 7 possible size outcomes. (b) Histogram with appropriate bins and corresponding probability, \( p_i \).](image)

In order to implement information entropy as a reliable measure of nanoparticle dispersity, three properties are required: (1) A linear relationship between entropy and population dispersity would facilitate interpretation and aid implementation into statistical optimization methods. (2) The entropy needs to be independent of the mean particle size. (3) The data needs to be discretized. Hence, we propose the following modification to the entropy calculation to produce the nanoparticle entropy, \( E \):

\[
E = e^{H_{\text{Bin Width}}}
\]

The response of the information entropy, \( H \), to increasing but equally probably outcomes (i.e., \( p_i = p_{\text{max}} \)) displays a logarithmic trend (Figure S1). Therefore, the exponential, a monotonic function, was added in eq 2. The resulting nanoparticle entropy, \( E \), increases linearly with dispersity. The nanoparticle entropy is independent of the mean particle size, another important characteristic. The same distribution will produce the same entropy regardless of the mean. Further information is found in the SI (Figures S1–S3). Please note that monodispersity criteria are based on the relative deviation from the mean particle size.\(^9\) To this end, a normalized entropy \( (E_n) \) can be obtained via dividing \( E \) by the mean. Since the nanoparticle entropy, \( E \), has the same units as the bin width, the normalized nanoparticle entropy, \( E_n \), follows as being dimensionless.

It is important to note that these entropy calculations require discrete data. While the size of nanoparticles is a continuous variable, both the imaging system and analysis method will impose a limit to the exactness of each measurement. This resolution becomes the bin width of a histogram, which effectively presents the nanoparticle distribution as a discrete data set. For an unbiased representation of the nanoparticle dispersity, the bin width must be included in the entropy calculation. The reasoning is as follows: a larger bin width will result in fewer bins and therefore a lower entropy; a smaller bin width for the same population will have a larger number of bins and a proportionally larger entropy. By including the bin width in the calculation of the entropy, this variability is avoided (see also Figures S4 and S5). Note that eq 2 assumes the use of a constant bin width.

Entropy depends on the sample size and asymptotically approaches the true value with increasing population. While several methods have been proposed to deal with this issue, we implement here the quadratic extrapolation by Strong et al. for its simplicity and low computational cost.\(^10\) This process relies on calculating \( E \) for the total population of \( M \) measurements and two subpopulations comprised of \( M/2 \) and \( M/4 \) measurements randomly selected from the main. This data is then fitted to eq 3, where \( x \) represents the sample size.

\[
E_M = E_{\text{true}} + \frac{b}{x} + \frac{a}{x^2} \tag{3}
\]

This method is powerful but requires sufficient data to adequately fit the quadratic. Figure 2 shows the results of the sample size correction for two different populations. For a given sample size, distributions with the characteristics of those shown in Figure 2a,c were randomly generated, and the nanoparticle entropy was calculated with and without sample size correction. This was repeated 100 times. Figure 2b,d shows the mean entropy with standard deviation as a function of the sample size for the respective populations. Without the sample size correction by quadratic extrapolation, at least 500 data points were required for population 1 and 900 for population 2 to achieve an entropy within 15% of the true value. With correction, these reduce to 100 and 150, respectively. On this account, we have developed a reliability index in the accompanying software (Matlab GUI and Excel Macro) to evaluate whether the sample size is sufficiently large for a reliable estimate of the entropy. We note that these sample size requirements are in line with sampling guidelines on conventional approaches.\(^11\) Further details about the sample size correction and the reliability index can be found in the SI.

In order to relate the normalized entropy \( E_n \) to established definitions of size uniformity, we have developed evaluation criteria for monodispersity based on definitions for dispersions provided by the National Institute of Standards and Technology (NIST) and guidelines used in nanocluster catalysis.\(^29,31,32\) The NIST requires that 90% of the particles must lie within \( \pm 5\% \) of the mean for a population to be considered monodisperse.\(^29\) In nanocluster analysis, a population is monodisperse if the standard deviation is \( \leq 5\% \) of the mean and near-monodisperse if it is \( \leq 15\% \) (\( \text{COV} = 0.05 \) and 0.15, respectively).\(^29,31,32\) It is important to note that the
guidelines from nanocluster analysis assume the population is normally distributed. A linear fit of the data in Figure S3 produces $E = 4.12\sigma$ and $E = \sigma\mu_{E\sigma}$, which corresponds to limits of 0.206 and 0.618 for monodispersity and near-monodispersity, respectively (Table 1).

The NIST standard presupposes no particular distribution. If we assume the distribution described in the NIST is Gaussian, the following must be true: COV $= 1.645$. Using the linear relationship between $E$ and $\sigma$, a limit of $E_{\sigma}$ follows with $0.125$. However, as the NIST guidelines do not specify a normal distribution, we evaluated the robustness of this limit for non-normal populations. A distribution was designed to maximize entropy in the limits of compliance with the NIST requirements; a schematic describing this shape is shown in Figure 3d. The range of 90% of the population is set by the NIST requirement, but the total range remains variant. We have therefore defined a variable $r$ as the ratio of range$_{90}$ and range$_{total}$, this is equal to the number of bins that lie in range$_{90}$ divided by the total number of bins. The relationship between $E_{\sigma}$ and $r$ is described in eq 4 and plotted in Figure 3. A detailed derivation of this equation can be found in the SI.

$$E_{\sigma} = \left(\frac{1}{9}\right)^{0.9} \left(\frac{1}{r} - 1\right)^{0.1}$$  \hspace{1cm} (4)

Figure 3a shows a 3D surface map of eq 4 for $n_{90}$ and $n_{10}$ in the range 10 to 105. (b) A 2D contour plot of eq 4; the contour lines denote normalized nanoparticle entropy at 0.1 intervals. The dashed lines indicate the cases where $r = 0.735$ and $r = 0.55$. The dotted line relates to the condition $T = 8 \times 10^4$. (c) Plot of the normalized nanoparticle entropy, $E_{\sigma}$ against $r$, $T = 8 \times 10^4$. (d) Schematic demonstrating the distribution that just meets the NIST criteria for monodispersity while producing the maximum entropy.

Table 1. Monodispersity Criteria: Normalized Nanoparticle Entropy Values for Monodisperse and Near-Monodisperse Populations

|                        | $E_{\sigma}$ |
|------------------------|-------------|
| highly monodisperse (NIST) | 0.125       |
| monodisperse (nanocluster) | 0.206       |
| near-monodisperse (nanocluster) | 0.618       |
distribution resulted in little change in $E_n$. We therefore recommend using the cutoff of $E_n = 0.125$, below which populations can be reliably considered as highly monodisperse.

## RESULTS AND DISCUSSION

In order to evaluate and benchmark this approach, six different data sets of oleylamine-capped gold nanoparticles were analyzed; see Figure 4. In each case, a representative transmission electron micrograph is displayed alongside the histogram obtained by image analysis. For each sample, the number of counts ($\approx 2000-5000$) is provided together with the mean size, standard deviation, and corresponding COV obtained by conventional analysis of the data sets. The absolute and normalized nanoparticle entropy for each nanoparticle population was obtained with the user-friendly macro (SI).

To check for normality, the apparent Gaussian distribution based on the mean and standard deviation calculated from the raw data was plotted alongside the histogram of the experimental data in Figure 4. While statistical tests for normality are abundant, these are often not suitable for large data sets. Experimental data will never produce a perfectly normal distribution at large sample sizes. Hence, the null hypothesis that the data is normal would be rejected in almost all cases. As noted previously, dispersity is the measure of the inhomogeneity of the size distribution rather than the breadth, which is correlated to the standard deviation. For populations which adhere to the Gaussian, standard deviation correlates directly with dispersity; as the population diverges from normality, this relationship breaks down and standard deviation becomes a less reliable measure of dispersity. For example, populations with apparent similar standard deviations, i.e., B and C, can display a significantly divergent degree of dispersity. Population B comprises a main population and a minor population; this small secondary population contains so few particles that it has little impact on the dispersity but does influence standard deviation. A similar behavior is observed between populations D and E. Populations F and E are comparable in dispersity but show a much larger disparity in standard deviation. Dispersity by definition is determined by the number of sizes observed and the relative quantities; in contrast, standard deviation depends on how these values are arranged around the mean. As population E consists of one broad distribution it results in a lower standard deviation than population F; the two peaks in population F shift the population density away from the mean producing a larger standard deviation for a similar range (see also Figure S11 in the SI).

![Figure 4](https://dx.doi.org/10.1021/acs.chemmater.0c00539)
Overall, discrepancies with the normal distribution (such as asymmetry or multiple modes, both commonly observed in nanoparticle populations) reduce the capacity of standard deviation to describe dispersity. Nanoparticle entropy, in contrast, remains a reliable metric irrespective of the type of distribution and reflects the significant differences between each population. This measure offers a clear pathway for optimization toward uniformity, i.e., by minimizing \( E \) or \( E_s \).

The populations presented in Figure 4 were taken from a larger study investigating the effects of experimental conditions on the dispersity of the resulting populations. Multiple studies highlighted the role of the reaction time, which should appear as a significant variable with effects such as size focusing. We have further analyzed data of a related study carried out in our lab. Therein, the measure of dispersity had a direct impact on its statistical significance. When using COV as a measure of dispersity, the reaction time was not identified as significant \( (p = 0.2404) \). In contrast, \( E_s \) identified the role of the reaction time on the homogeneity of the particle size \( (p = 0.0009) \). The large difference between these two \( p \)-values is consistent with the imprecision of COV and other like methods when measuring dispersity of non-normal populations. By not using an exact measure of dispersity, the cumulative error obscures important synthetic variables, here the role of the reaction time on the particle size distribution. This example highlights the importance of using an appropriate metric.

Please note that the method can be used on any data set that provides a histogram of the nanoparticle population, e.g., nanoparticle tracking analysis, disc centrifugation analysis, analytical ultracentrifugation, small-angle X-ray scattering, or dynamic light scattering. For the latter ensemble-based techniques, the data can be presented in a suitable format by approximating the number of measurements. Please note that, as the number of measurements usually exceed \( 10^5 \) particles, an order of magnitude estimation will be more than sufficient with little to no impact on the final calculated nanoparticle entropy.

**CONCLUSION**

To conclude, we propose the use of nanoparticle entropy as a reliable measure to evaluate dispersity in nanoparticle populations. This approach allows any type of distribution to be described, irrespective of being mono-, poly-, or heterodisperse. We envision this approach to be particularly useful for optimization protocols that are targeted toward achieving size uniformity. In particular, nanoparticle entropy represents a reliable descriptor in automated synthetic procedures leveraging on advanced statistical tools, including design of experiment and machine learning.

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.0c00539. The MATLAB code as well as a user-friendly Macro for Microsoft Excel (Windows and MacOS) are available via https://github.com/adrena-lab/Nanoparticle_Entropy.

Further details regarding the response of entropy as a function of dispersity; measurement resolution and bin width; sample size correction; further details on the determination of the NIST monodispersity criteria; and statistics on the experimental data (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

Stefan Guldin – University College London, Department of Chemical Engineering, Torrington Place, London WC1E 6BT, United Kingdom; orcid.org/0000-0002-4413-5527; Email: s.guldin@ucl.ac.uk

**Author**

Niamh Mac Fhionnlaioch – University College London, Department of Chemical Engineering, Torrington Place, London WC1E 6BT, United Kingdom; orcid.org/0000-0001-8459-9994

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acs.chemmater.0c00539

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

N.M.F. acknowledges funding by the EPSRC under a Doctoral Training Partnership (EP/M507970/1). The authors are grateful to Dr. Federico Galvanin and Dr. Marco Quaglio for guidance and feedback.

**REFERENCES**

(1) Pan, Y.; Neuss, S.; Leifert, A.; Fischler, M.; Wen, F.; Simon, U.; Schmidt, G.; Brandau, W.; Jahnen-Dechent, W. Size-Dependent Cytoxicity of Gold Nanoparticles. Small 2007, 3, 1941–1949.

(2) Zhou, X.; Xu, W.; Liu, G.; Panda, D.; Chen, P. Size-Dependent Catalytic Activity and Dynamics of Gold Nanoparticles at the Single-Molecule Level. J. Am. Chem. Soc. 2010, 132, 138–146.

(3) Saha, K.; Agasti, S. S.; Kim, C.; Li, X.; Rotello, V. M. Gold Nanoparticles in Chemical and Biological Sensing. Chem. Rev. 2012, 112, 2739–2779.

(4) MacDonald, T. J.; Wu, K.; Sehmi, S. K.; Noimark, S.; Peveler, W. J.; Du Toit, H.; Voelcker, N. H.; Allan, E.; MacRobert, A. J.; Gavrilidis, A.; Parkin, I. P. Thiol-Capped Gold Nanoparticles Swell-Encapsulated into Polyurethane as Powerful Antibacterial Surfaces under Dark and Light Conditions. Sci. Rep. 2016, 6, 39272.

(5) Loyanchan, C. N.; Soleimany, A. P.; Dudani, J. S.; Lin, Y.; Najer, A.; Bekdemir, A.; Chen, Q.; Bhatia, S. N.; Stevens, M. M. Renal clearable catalytic gold nanoclusters for in vivo disease monitoring. Nanotechnol. 2019, 14, 883–890.

(6) Gaumet, M.; Vargas, A.; Gurny, R.; Delie, F. Nanoparticles for drug delivery: The need for precision in reporting particle size parameters. Eur. J. Pharm. Biopharm. 2008, 69, 1–9.

(7) Baalousha, M.; Lead, J. R. Nanoparticle dispersity in toxicology. Nat. Nanotechnol. 2013, 8, 308–309.

(8) Murphy, C. J.; Buriak, J. M. Best Practices for the Reporting of Colloidal Inorganic Nanomaterials. Chem. Mater. 2015, 27, 4911–4913.

(9) Modena, M. M.; Rühle, B.; Burg, T. P.; Wuttke, S. Nanoparticle Characterization: What to Measure? Adv. Mater. 2019, 1901556.

(10) Muzzio, M.; Li, J.; Yin, Z.; Delahunty, I. M.; Xie, J.; Sun, S. Monodisperse nanoparticles for catalysis and nanomedicine. Nano- scale 2019, 11, 18946–18967.

(11) Wolf, M.; Fischer, N.; Claeyss, M. Surfactant-free synthesis of monodisperse cobalt oxide nanoparticles of tunable size and oxidation state developed by factorial design. Mater. Chem. Phys. 2018, 213, 305–312.

(12) Quintero-Quiroz, C.; Acevedo, N.; Zapata-Giraldo, J.; Botero, L. E.; Quintero, J.; Zárate-Trivinó, D.; Saldarrriaga, J.; Pérez, V. Z. Optimization of silver nanoparticle synthesis by chemical reduction and evaluation of its antimicrobial and toxic activity. Biomaterials Research 2019, 23, 27.
(13) Keijok, W. J.; Pereira, R. H. A.; Alvarez, L. A. C.; Prado, A. R.; da Silva, A. R.; Ribeiro, J.; de Oliveira, J. P.; Guimarães, M. C. C. Controlled biosynthesis of gold nanoparticles with Coffea arabica using factorial design. Sci. Rep. 2019, 9, 16019.

(14) Mora-Tamez, L.; Barim, G.; Downes, C.; Williamson, E. M.; Habas, S. E.; Brutchez, R. L. Controlled Design of Phase- and Size-Tunable Monodisperse Ni2P Nanoparticles in a Phosphonium-Based Ionic Liquid through Response Surface Methodology. Chem. Mater. 2019, 31, 1552–1560.

(15) Musumeci, T.; Bonaccorso, A.; Carbone, C.; Russo, G.; Pappalardo, F.; Puglisi, G. Design and optimization of PEGylated nanoparticles intended for Berberine Chloride delivery. J. Drug Delivery Sci. Technol. 2019, 52, 521–530.

(16) Yu, A.; Standish, N. A study of particle size distributions. Powder Technol. 1990, 62, 101–118.

(17) Fieller, N. R. J.; Flennley, E. C.; Olbricht, W. Statistics of Particle Size Data. Applied Statistics 1992, 41, 127.

(18) Merkus, H. G. Particle Size Measurements, 1st ed.; Springer Netherlands: Dordrecht, 2009; Particle S, pp 13–42.

(19) Scheibelhofer, O.; Besenhard, M. O.; Piller, M.; Khinast, J. G. Comparing particle size distributions of an arbitrary shape. Powder Technol. 2016, 294, 134–145.

(20) Defante, A. P.; Vreeland, W. N.; Benkstein, K. D.; Ripple, D. C. Using Image Attributes to Assure Accurate Particle Size and Count Using Nanoparticle Tracking Analysis. J. Pharm. Sci. 2018, 107, 1383–1391.

(21) Fathi, P.; Knox, H. J.; Sar, D.; Tripathi, I.; Ostadbhossein, F.; Misra, S. K.; Esch, M. B.; Chan, J.; Pan, D. Biodegradable Biliverdin Nanoparticles for Efficient Photoacoustic Imaging. ACS Nano 2019, 13, 7690–7704.

(22) Meijerink, M. J.; de Jong, K. P.; Zečević, J. Growth of Supported Gold Nanoparticles in Aqueous Phase Studied by in Situ Transmission Electron Microscopy. J. Phys. Chem. C 2020, 124, 2202–2212.

(23) Shannon, C. E. A Mathematical Theory of Communication. Bell Syst. Tech. J. 1948, 27, 379–423.

(24) Karmeshu; Pal, N. Studies in Fuzziness and Soft Computing. In Entropy Measures, Maximum Entropy Principle and Emerging Applications; Karmeshu, Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2003; Vol. 119, Chapter 1, pp 1–54.

(25) MacArthur, R. H. Patterns of Species Diversity. Biological Reviews 1965, 40, 510–533.

(26) Chao, A.; Jost, L.; Hsieh, T. C.; Ma, K. H.; Sherwin, W. B.; Rollins, L. A. Expected Shannon Entropy and Shannon Differentiation between Subpopulations for Neutral Genes under the Finite Island Model. PLoS One 2015, 10, No. e0125471.

(27) Bonchev, D.; Kamenski, D.; Kamenska, V. Symmetry and information content of chemical structures. Bull. Math. Biophys. 1976, 38, 119–133.

(28) Zhou, R.; Cai, R.; Tong, G. Applications of Entropy in Finance: A Review. Entropy 2013, 15, 4909–4931.

(29) Hackley, V. A.; Ferraris, C. F. NIST Recommended Practice Guide: The Use of Nomenclature in Dispersion Science and Technology — NIST; Special Publication (NIST SP) 960-3; 2001; pp 1–64.

(30) Strong, S. P.; Koberle, R.; de Ruyter van Steveninck, R. R.; Bialek, W. Entropy and Information in Neural Spike Trains. Phys. Rev. Lett. 1998, 80, 197–200.

(31) Aiken, J. D., III; Lin, Y.; Finke, R. G. A perspective on nanocluster catalysis: polyoxoanion and (n-C4H9)4N+ stabilized Ir(0)−300 nanocluster soluble heterogeneous catalysts. J. Mol. Catal. A: Chem. 1996, 114, 29–51.

(32) Sun, S.; Murray, C. B.; Weller, D.; Folks, L.; Moser, A. Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal superlattices. Science 2000, 287, 1989–92.

(33) D’agostino, R. B.; Belanger, A.; D’agostino, R. B. A Suggestion for Using Powerful and Informative Tests of Normality. Am. Stat. 1990, 44, 316–321.