

RESEARCH ARTICLE

# The Power of Heterogeneity: Parameter Relationships from Distributions

Magnus Röding<sup>1,2,5\*</sup>, Siobhan J. Bradley<sup>2,4</sup>, Nathan H. Williamson<sup>2</sup>, Melissa R. Dewi<sup>2,3</sup>, Thomas Nann<sup>2,3,4</sup>, Magnus Nydén<sup>5</sup>

**1** SP Food and Bioscience, Soft Materials Science, Göteborg, Sweden, **2** Future Industries Institute, University of South Australia, Adelaide, Australia, **3** ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, Ian Wark Research Institute, University of South Australia, Adelaide, Australia, **4** MacDiarmid Institute, Victoria University of Wellington, Wellington, New Zealand, **5** School of Energy and Resources, UCL Australia, University College London, Adelaide, Australia

\* [magnus.rodning@sp.se](mailto:magnus.rodning@sp.se)



**OPEN ACCESS**

**Citation:** Röding M, Bradley SJ, Williamson NH, Dewi MR, Nann T, Nydén M (2016) The Power of Heterogeneity: Parameter Relationships from Distributions. PLoS ONE 11(5): e0155718. doi:10.1371/journal.pone.0155718

**Editor:** Alessandro Esposito, University of Cambridge, UNITED KINGDOM

**Received:** August 27, 2015

**Accepted:** May 3, 2016

**Published:** May 16, 2016

**Copyright:** © 2016 Röding et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was funded by the South Australian Government Premier's Research and Industry Grant project "A Systems Approach to Surface Science" and partly by the Australian Research Council Centre of Excellence in Convergent Bio-Nano Science and Technology (project number CE140100036). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Complex scientific data is becoming the norm, many disciplines are growing immensely data-rich, and higher-dimensional measurements are performed to resolve complex relationships between parameters. Inherently multi-dimensional measurements can directly provide information on both the distributions of individual parameters and the relationships between them, such as in nuclear magnetic resonance and optical spectroscopy. However, when data originates from different measurements and comes in different forms, resolving parameter relationships is a matter of data analysis rather than experiment. We present a method for resolving relationships between parameters that are distributed individually and also correlated. In two case studies, we model the relationships between diameter and luminescence properties of quantum dots and the relationship between molecular weight and diffusion coefficient for polymers. Although it is expected that resolving complicated correlated relationships require inherently multi-dimensional measurements, our method constitutes a useful contribution to the modelling of quantitative relationships between correlated parameters and measurements. We emphasise the general applicability of the method in fields where heterogeneity and complex distributions of parameters are obstacles to scientific insight.

## Introduction

Increasing the dimensions of an experiment allows the measurement of dependence between parameters. For example, introducing multi-dimensional nuclear magnetic resonance (NMR) has opened new possibilities for studying heterogeneous structures and complex phenomena by correlating different parameters describing transport properties and identifying different populations based on diffusion and relaxation properties [1, 2]. In optical (electronic) spectroscopy in the infrared, visible, and ultraviolet ranges, obtaining 2D spectra that describe correlations of excitation and detection wavelengths recently enabled the mapping of energy transfer

**Competing Interests:** The authors have declared that no competing interests exist.

pathways in the light-harvesting mechanism of certain bacterial species [3–5]. The key notion for both techniques is that a *joint* probability distribution (or ‘spectrum’ or ‘map’) of different parameters and their relationships can be estimated, instead of just the *marginal* (or lower-dimensional) distributions for individual parameters.

Multi-dimensional *intra*-modality measurements—within the confines of a single technique—inherently provide probabilistic relationships between parameters. More often than not, though, different parameters of systems are studied using different techniques, i.e. a multi-dimensional *inter*-modality measurement; revealing parameter relationships then becomes a matter of data analysis post-experiment, rather than a matter of experiment per se. Unfortunately, individual parameters in almost all real systems are distributed, and heterogeneity is a major obstacle to data interpretation. One common approach to this problem is to rely on purification to minimise sample heterogeneity, estimating relationships using standard methods such as regression by least-squares fitting. Altering samples is typically not viable in the industrial processing of heterogeneous raw materials, and this can be a substantial hurdle even in controlled laboratory settings.

This dilemma has prompted us to move to a setting where experimental input and output originate from different measurement techniques and comprise distributions rather than single values. Combining information from different sources, often called data integration or data fusion [6], is broadly addressed in computer science, machine learning, and mathematics. Modelling probabilistic relationships when data is obtained as distributions of individual parameters is almost unexplored; the closest work has discussed ‘distribution to distribution regression’ in a different context [7].

Here we develop a widely applicable general-purpose statistical method for estimating probabilistic relationships from distributions of data for individual parameters. Critically, these distributions can be represented by different types of data, i.e. statistical samples (sets of values), decay curves, spectra, function curves, and histograms, and therefore involve solving inverse problems.

## Results and Discussion

Consider the probabilistic relationship between an input parameter  $x$  and an output parameter  $y$  by modelling the joint distribution describing the dependence structure of  $x$  and  $y$ , based on measurements from  $K$  samples. The data from the  $k$ th sample represents marginal distributions  $p_{x;k}(x)$  and  $p_{y;k}(y)$  (directly as a statistical sample, or indirectly through an inverse problem) from the joint distribution for the  $k$ th sample,  $p_{x, y;k}(x, y)$ . Generally, nothing can be said about the dependence structure of  $x$  and  $y$  from a single measurement alone: finding a joint distribution from marginal distributions is by itself an inverse problem that requires assumptions or a priori knowledge. However, assuming that the conditional distribution of  $y$  given  $x$  is common to all samples, the joint distribution of  $x$  and  $y$  for the  $k$ th sample can be written  $p_{x, y;k}(x, y) = p_{y|x}(x, y)p_{x;k}(x)$ . Hence,

$$p_{y;k}(y) = \int_0^{\infty} p_{y|x}(x, y)p_{x;k}(x) dx. \tag{1}$$

A ‘link’ between the measurements for different samples is provided through the conditional distribution  $p_{y|x}(x, y)$ , interpreted physically as the distribution of  $y$  caused by unobserved parameters of the system other than  $x$ . That it is independent of sample index  $k$  corresponds to assuming that sorting into  $K$  samples is conducted solely on the basis of the value of  $x$ . The conditional distribution  $p_{y|x}(x, y)$  is estimated indirectly by fitting the distributions  $p_{y;k}(y)$  to the measurements, yielding an estimate of the probabilistic relationship between  $x$  and  $y$ .

Thus, a space of distributions of  $x$  is mapped onto a space of distributions of  $y$ , in the process of modelling the relationship. The distribution models may be parametric as well as nonparametric.

First, we study probabilistic relationships between material parameters for colloidal semiconductor quantum dots (QDs), which are luminescent nanoparticles with applications in imaging [8–10], catalysis [11], photovoltaics [12–14], and many more fields. QDs comprise a prototypical complex material with properties dependent on the interaction between multiple parameters. In particular, the emission spectral profile and excited-state decay dynamics and their relation to size are vital for understanding the electro-optic and catalytic properties of these materials [15]. Additionally, QD size polydispersity is often inherent to synthesis and further purification to reduce the polydispersity is often impractical. A set of four batches of standard ZnS-coated CdSe QDs are synthesised and diameter distributions estimated from transmission electron microscopy (TEM) image data. Steady-state emission spectra of the QDs are acquired and interpreted as distributions of wavelengths  $p_{\lambda;k}(\lambda)$  convolved with a Lorentzian line-broadening function,  $\gamma/(\pi((\lambda - \lambda')^2 + \gamma^2))$ . The model for the  $k$ th spectrum is therefore of the form

$$I_k(\lambda) = I_{0;k} \int_0^\infty p_{\lambda;k}(\lambda') \frac{\gamma/\pi}{(\lambda - \lambda')^2 + \gamma^2} d\lambda'. \tag{2}$$

The distribution  $p_{\lambda;k}(\lambda)$  is the marginal distribution of  $\lambda$  from the joint distribution

$$p_{d,\lambda;k}(d, \lambda) = p_{\lambda|d}(d, \lambda) p_{d;k}(d). \tag{3}$$

This factoring corresponds to assuming that the QDs are sorted only according to size, and that other variation for a given diameter  $d$  is the same regardless of the sample. The distribution  $p_{\lambda|d}(d, \lambda)$  constitutes a summary of the influence of all system parameters other than  $d$ . If it is degenerate,  $p_{\lambda|d}(d, \lambda) = \delta(\lambda - \lambda_0)$ ,  $\lambda$  is completely determined by  $d$ , and thus there is no other relevant parameter in the relationship. Fig 1 shows the spectra, the model fits, and the estimated joint distributions of diameters and emission wavelengths.

We now acquire fluorescence lifetime distributions of the QDs and interpret them as superpositions of exponential lifetime distributions, governed by a ‘characteristic’ (average) lifetime distribution for the  $k$ th sample,  $p_{\tau;k}(\tau)$ , such that

$$p_{t,k}(t) = \int_0^\infty p_{\tau;k}(\tau) \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) d\tau. \tag{4}$$

The distribution  $p_{\tau;k}(\tau)$  is the marginal distribution of  $\tau$  coming from the joint distribution

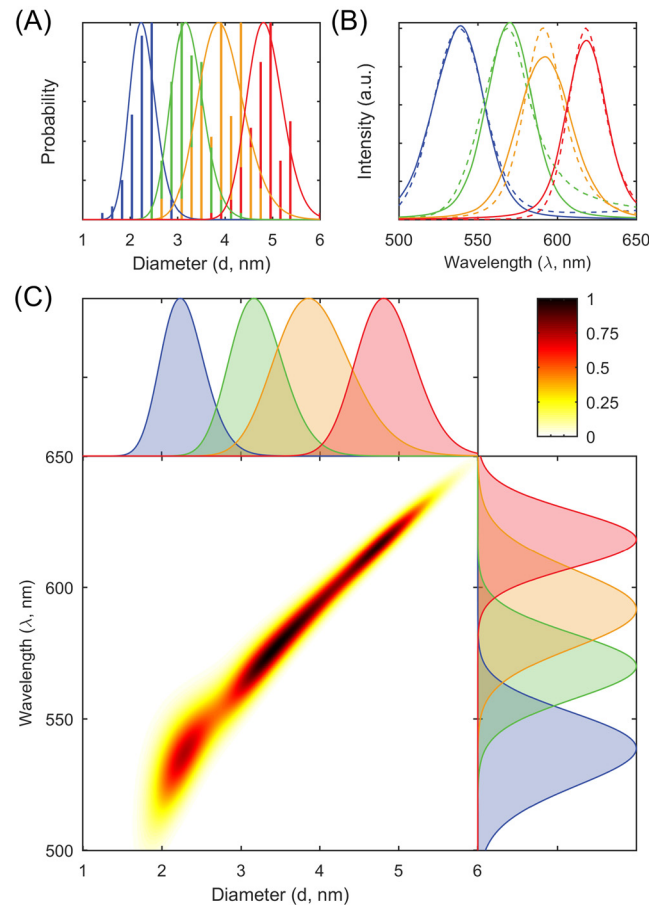
$$p_{d,\tau;k}(d, \tau) = p_{\tau|d}(d, \tau) p_{d;k}(d). \tag{5}$$

Otherwise, the analysis is identical to the previous one for wavelengths. Fig 2 shows the fluorescence lifetime distributions, the model fits, and the estimated joint distributions of diameters and characteristic lifetimes. The results show that the larger particles have narrower lifetime distributions. This may be due to the presence of fewer defects related to a lower surface-to-volume ratio.

Second, we study the probabilistic relationship between molecular weight  $M$  and diffusion coefficient  $D$  for polymers. A well-known empirical scaling law for linear polymer chains states that

$$D(M) = KM^{-\nu} \Leftrightarrow M(D) = K^{1/\nu} D^{-1/\nu}, \tag{6}$$

where  $\nu \sim 0.5 - 0.8$  for uncharged polymers in the dilute regime, depending on solvent quality and temperature [16]. The parameter  $\nu$ , analogous to the Flory exponent [17], is a measure of

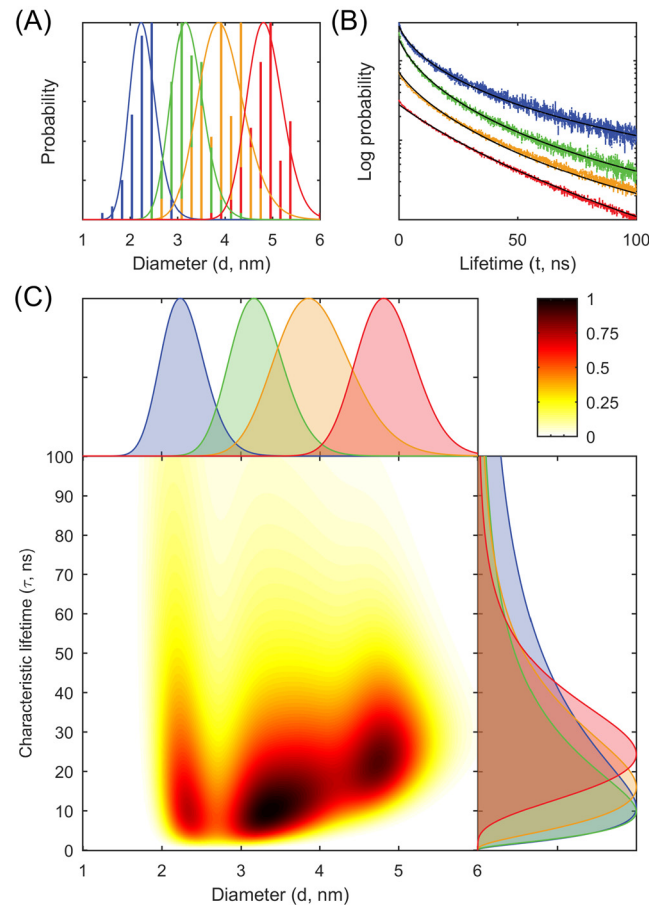


**Fig 1. Correlating diameters to emission wavelengths.** (A) Diameter histograms ( $n = 95, 127, 88,$  and  $106$ ) for the four samples (blue, green, yellow, and red) extracted from transmission electron microscopy and the model fits (solid lines). (B) Steady-state emission spectra (dashed lines) and the model fits (solid lines) for the four samples (blue, green, yellow, and red). (C) Estimated joint distributions of diameters and wavelengths with marginal diameter distributions at the top and marginal wavelength distributions at the right.

doi:10.1371/journal.pone.0155718.g001

how polymers spread in space [18, 19]. Knowledge of both parameters  $K$  and  $\nu$  allows determination of molecular weight distributions from diffusion coefficient distribution measurements [20]. Polymers comprise a prototypical example of the case where samples are often purified to minimise heterogeneity, typically estimating the scaling law parameters  $K$  and  $\nu$  by taking logarithms and using linear least-squares on a set of many samples [21]. Considering a polydisperse sample with marginal distributions  $p(M)$  and  $p(D)$ , the scaling law relationship can be seen as a degenerate joint distribution with non-zero density only on the line described by Eq 6 because no other parameter influences the relationship, hence it is a special case of the more general example above. We use the fact that if  $p(M)$  is a lognormal distribution with parameter  $(\mu_M, \sigma_M)$ ,  $p(D)$  is also a lognormal distribution with parameter  $(\mu_D, \sigma_D)$ . On a polydisperse polystyrene sample, we measure the molecular weight distribution by gel permeation chromatography (GPC), obtained as a distribution over  $\log_{10} M$ . We measure the diffusion coefficient distribution of the polystyrene, diluted in deuterated chloroform ( $\text{CDCl}_3$ ), by nuclear magnetic resonance (NMR), obtaining a signal attenuation described by the Stejskal-Tanner equation [22]

$$I(b) = I_0 \int_0^\infty p(D) \exp(-bD) dD \quad (7)$$



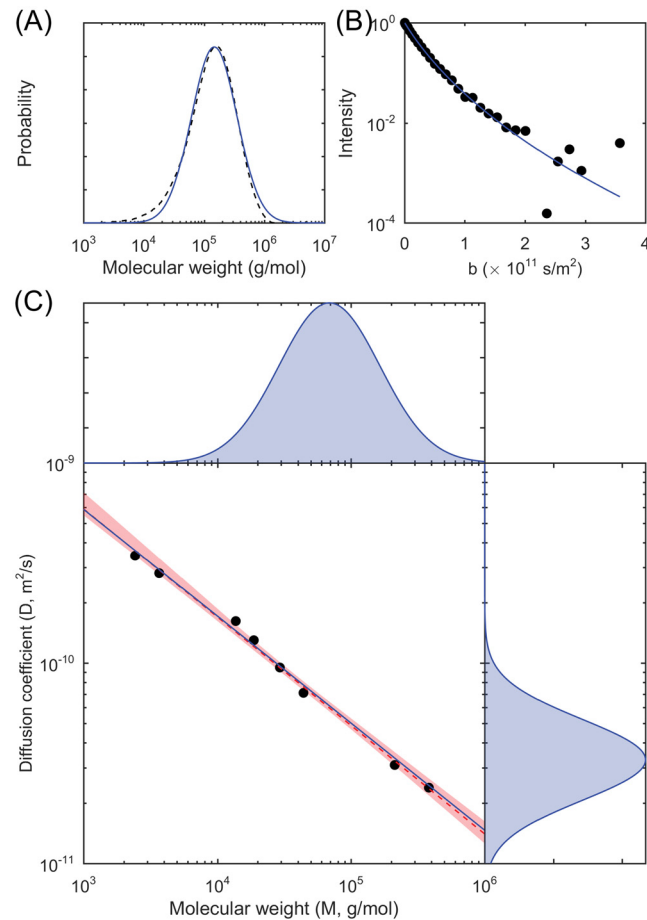
**Fig 2. Correlating diameters to characteristic fluorescence lifetimes.** (A) Diameter histograms ( $n = 95, 127, 88,$  and  $106$ ) for the four samples (blue, green, yellow, and red) extracted from transmission electron microscopy and the model fits (solid lines). (B) Fluorescence lifetime measurements and the model fits (black solid lines, overlaid on top of the experimental data) for the four samples (blue, green, yellow, and red, vertical rescaling has been performed to avoid occlusion). (C) Estimated joint distributions of diameters and characteristic lifetimes with marginal diameter distributions at the top and marginal characteristic lifetime distributions at the right.

doi:10.1371/journal.pone.0155718.g002

for a lognormal distribution  $p(D)$  with parameters  $(\mu_D, \sigma_D)$ . From these parameters, the scaling law parameters  $K$  and  $\nu$  can be obtained by

$$\begin{aligned} \nu &= \frac{\sigma_D}{\sigma_M} \\ K &= \exp(\mu_D + \nu\mu_M). \end{aligned} \tag{8}$$

As a validation, we use the reported values of  $M_w$  from the manufacturer for eight well-defined monodisperse polystyrene standards and the corresponding values of  $\langle D \rangle$ , measured by NMR for dilute amounts of the samples in  $CDCl_3$ , and perform the conventional linear least-squares fitting to estimate  $K$  and  $\nu$ . Fig 3 shows the molecular weight distribution data, the diffusion coefficient distribution data, the model fits, and the estimated scaling law relationship. For dilute polystyrene in  $CDCl_3$  at ambient conditions, we obtain  $K = 2.47 \times 10^{-8}$  and  $\nu = 0.541$  with the conventional method. Using the proposed distribution-based method, we perform experiments in triplicate and obtain  $K = 2.35 \times 10^{-8}, 2.67 \times 10^{-8},$  and  $2.16 \times 10^{-8}$ , and  $\nu = 0.534, 0.544,$  and  $0.526$ . These results are not significantly different from the one obtained



**Fig 3. Correlating molecular weights to diffusion coefficients.** (A) Molecular weight distribution measured by gel permeation chromatography (dashed black line) and the model fit (solid blue line). (B) Nuclear magnetic resonance attenuation data and the model fit. (C) Estimated joint distribution, degenerating to a distribution on a line, of molecular weight and diffusion coefficient, with the data points for the monodisperse samples (black disks) and the relationship estimated with the conventional method from these samples (dashed red line, with error bounds), relationship estimated with the proposed distribution-based method (solid blue line), and with marginal molecular weight distribution at the top and marginal diffusion coefficient distribution at the right for the polydisperse sample.

doi:10.1371/journal.pone.0155718.g003

with the conventional method based on a 95% nonparametric confidence interval of the latter. That  $K$  and  $\nu$  can be determined from a single sample has been suggested before [23], but is here performed for the first time.

We emphasise that instead of needing to perform a painstakingly large number of measurements on monodisperse samples (and possibly needing to perform purification steps to obtain those samples), we can obtain a very similar result by different and simpler means, based on performing a single measurement on a polydisperse sample in this case when the functional form of the relationship is known.

### Conclusions

We demonstrate a general-purpose mathematical method for estimating probabilistic relationships from measured distributions of individual parameters. The distributions can be represented directly by statistical samples, or indirectly through an inverse problem by decay curves, spectra,

function curves, and histograms. The result is an estimate of a joint probability distribution of parameters that describes relationships and correlations. In the applications to quantum dots, measurements from multiple heterogeneous samples are utilized, allowing for joint probability distributions (Figs 1 and 2) for which the effects of diameter are observed through the dependence of both the mean and the variance of the conditional distribution on diameter. In the application to polymers, using measurements on a single polydisperse polymer sample, the parameters  $K$  and  $\nu$  of the scaling law relationship in Eq 6 are estimated, and this gives rise to a joint probability distribution (Fig 3) (with conditional distributions being in fact delta functions).

In a world ever more data-rich, with a plethora of measurement techniques and sensors used in the minerals, pharmaceutical, food, and chemical industries, as well as in fundamental science, we urgently need to improve the ways in which we analyse the complex data sets that are becoming the norm. This trend will only continue, in particular because natural raw materials need to be used to their full potential to meet future sustainability and cost demands—materials that are naturally heterogeneous, complex, and challenging to master. We expect our general-purpose mathematical method to be applicable in many fields where heterogeneity and complex distributions of parameters are obstacles to scientific insight. Single values of parameters can be replaced by distributions, taking advantage of the heterogeneity and letting it work for us, not against us. Naturally, multi-dimensional experiments deliver a wealth of information which is literally impossible to access with lower-dimensional experiments. Though more complicated parameter relationships, involving e.g. non-monotonicity, discontinuity, or multimodal conditional distributions, cannot be resolved with our method, we expect the method to be useful for the understanding of many heterogeneous systems.

## Materials and Methods

### Preparation of quantum dots

Four samples of ZnS-coated CdSe core-shell quantum dots (QDs) with different nominal diameters in the range  $d \approx 2 - 5$  nm are synthesized using a previously published method [24] and dissolved in toluene.

### Transmission electron microscopy measurements

Diameters of QDs are estimated using transmission electron microscopy (TEM). The analysis is performed on a JEOL 2100F microscope operated at 200 kV. All experiments are carried out at room temperature. The spatial resolutions (physical pixel sizes) are  $\Delta x = 0.05 - 0.17$  nm for the image data. Particle contours ( $n = 95, 127, 88,$  and  $106$ ) for the four samples are identified manually from raw TEM images.

### Luminescence measurements

Steady-state emission spectra of the QDs are acquired in the range 450–650 nm with resolution  $\Delta\lambda = 1$  nm using an Edinburgh Photonics FLS 980 time-resolved photoluminescence spectrometer. At each respective emission intensity maxima, fluorescence lifetime measurements by time-correlated single photon counting (TCSPC) are performed using the same instrument. Excitation lifetimes are recorded in a 500 ns range using 8192 channels.

### Nuclear magnetic resonance measurements

Polydisperse polystyrene (PS) (190,000  $M_w$ , Sigma-Aldrich) is mixed to 0.02% w/w in  $CDCl_3$  (99.8 atom% deuterium, Sigma-Aldrich). Monodisperse PS standards (2,430  $M_w$ , 3,680  $M_w$ , 13,700  $M_w$ , 18,700  $M_w$ , 29,300  $M_w$ , 44,000  $M_w$ , 212,400  $M_w$ , and 382,100  $M_w$ , Sigma-Aldrich)

are each mixed to 0.1% w/w, also in CDCl<sub>3</sub>. These concentrations are chosen in order to be within the dilute polymer regime (18) and to avoid microscopic averaging effects [25]. NMR tubes are filled with the PS solutions and flame-sealed to avoid convection due to solvent evaporation. Pulsed Gradient Stimulated Echo (PGSTE) measurements of <sup>1</sup>H self-diffusion [26] are performed at ambient conditions using a Bruker Avance III HD NMR spectrometer with an Ascend 600 MHz superconducting magnet, a micro5 probe, and a diff30 gradient coil. The PGSTE parameters for all measurements are repetition time TR = 10 s, gradient pulse duration δ = 1.576 ms, observation time Δ = 50 ms, and τ = 2.6 ms. The PGSTE measurements on the 0.02% w/w 190,000 M<sub>w</sub> polydisperse PS sample, performed in triplicate, use 192 averages and 32 linearly spaced gradient steps with maximum gradient value 10 T/m. The PGSTE measurements on the PS standard samples use 16 averages and 16 linearly spaced gradient steps with maximum gradient values chosen to attenuate 30% of the PS signals. The spectrally resolved PS signal attenuations are analysed using the Stejskal-Tanner equation, Eq 7, with attenuation factor

$$b = (\gamma g \delta)^2 \frac{4}{\pi^2} \left( \Delta - \frac{\delta}{4} \right), \tag{9}$$

for a sinusoidal gradient pulse shape.

### Gel permeation chromatography measurements

Gel permeation chromatography (GPC) molecular weight distribution measurements of the 190,000 M<sub>w</sub> polydisperse PS are performed on an Agilent PL-GPC 220 system flowing trichlorobenzene at 1 ml/min through three PLgel 10 μm mixed-B columns at 150°C. Measurements are performed in triplicate on 1.5 mg/ml samples using a 200 μl injection volume and conventional calibration with Agilent EasiVial PS-H standards. The data is obtained as 400–500 points equidistant in log<sub>10</sub> M.

### Data analysis for quantum dots

All particle contours identified from the TEM images are near-circular, and an equivalent diameter is computed by

$$d = \left( \frac{4A}{\pi} \right)^{1/2} = 2 \left( \frac{N}{\pi} \right)^{1/2} \Delta x, \tag{10}$$

where A is area and N is number of pixels within the contour. We make parametric model assumptions and fit lognormal distributions,

$$p_{d,k}(d) = \frac{1}{d \sigma_{d,k} \sqrt{2\pi}} \exp \left( - \frac{(\log(d) - \mu_{d,k})^2}{2\sigma_{d,k}^2} \right), \tag{11}$$

to the diameters using maximum likelihood [27], resulting in parameter estimates

$$\begin{aligned} \mu_{d,k} &= \frac{1}{n} \sum_i \log(d_i) \\ \sigma_{d,k} &= \frac{1}{n} \sum_i (\log(d_i) - \mu_{d,k})^2. \end{aligned} \tag{12}$$

Steady-state emission spectra are interpreted as probability distribution of wavelengths convolved with a Lorentzian curve (Cauchy distribution with parameter γ, corresponding to a FWHM of 2γ), accounting for line broadening. All spectra are fit simultaneously using



nonlinear least-squares, thus simultaneously estimating  $p_{\lambda|d}(d, \lambda)$  and  $p_{\lambda;k}(\lambda)$  for fixed  $p_{d;k}(d)$  by minimizing the global sum of squares

$$SS = \sum_k \sum_l (I_k(\lambda_l) - I_k^{\text{exp}}(\lambda_l))^2, \tag{13}$$

summed over all spectra and all wavelengths. For fluorescence lifetime measurements by TCSPC, time zero is determined by finding the channel with the highest photon count. The lifetime distributions are interpreted as superpositions of exponential distributions,

$$p_{t,k}(t) = \int_0^\infty p_{\tau,k}(\tau) \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) d\tau. \tag{14}$$

governed by ‘characteristic’ (average) lifetime distributions  $p_{\tau,k}(\tau)$ . Data is acquired in discrete time, as a histogram, and fitting of the discrete-time counterpart of Eq 14 with an additional baseline (dark photon count) probability is performed using maximum likelihood and ‘discretized’ channel probabilities, corresponding to the histogram binning,  $q_{k,l}$  for bin  $l$ . For further details on fitting refer to Rödning (2014) [28]. Because the photon counts differ between lifetime measurements (between  $N = 2.6 \times 10^5 - 2.1 \times 10^6$ ), weighted maximum likelihood is employed. All lifetime distributions are fit simultaneously using weighted maximum likelihood, thus simultaneously estimating  $p_{\tau|d}(d, \tau)$  and  $p_{\tau;k}(\tau)$  for fixed  $p_{d;k}(d)$  by maximizing the weighted global loglikelihood function

$$\log L = \sum_k \sum_l \frac{n_{k,l}}{N_k} \log q_{k,l}, \tag{15}$$

summed over all lifetime distributions and all bins, and where  $n_{k,l}$  is the number of photons in bin  $l$  of distribution  $k$ , and  $N_k$  is the total number of photons in distribution  $k$ . Integrals in  $(d, \lambda)$  and  $(d, \tau)$  space are computed using monte carlo integration with  $n_{\text{mc}} = 10^4$  quasi-random samples [29]. As a model for  $p_{\lambda|d}(d, \lambda)$  we choose a lognormal distribution with diameter-dependent mean  $m_\lambda(d)$  and standard deviation  $s_\lambda(d)$ . We let both  $m_\lambda(d)$  and  $s_\lambda(d)$  be power-law type functions,

$$m_\lambda(d) = m_1 + m_2 d^{m_3} \tag{16}$$

and

$$s_\lambda(d) = s_1 + s_2 d^{s_3}. \tag{17}$$

Data analysis is performed using Matlab R2015a (Mathworks, Natick, MA, US).

### Data analysis for polymers

Molecular weight distributions measured by GPC are fit using a normal distribution in  $\log_{10} M$  space with parameters  $(\mu_M/\log 10, \sigma_M/\log 10)$ , using nonlinear least-squares and minimizing

$$SS = \sum_i \left( p_{\text{GPC}}(\log_{10} M_i) - p_{\text{Normal}}\left(\log_{10} M_i; \frac{\mu_M}{\log 10}, \frac{\sigma_M}{\log 10}\right) \right)^2. \tag{18}$$

By a well-known scaling property of the lognormal distribution family, the scaling law in Eq (6) implies that  $D$  is also lognormal distributed with parameters

$$\begin{aligned} \mu_D &= \log K - v\mu_M \\ \sigma_D &= v\sigma_M. \end{aligned} \tag{19}$$

This derivation is fairly straightforward by use of the standard change-of-variables technique for probability distributions. Hence, a lognormal distribution model for  $p(D)$  is used in the Stejskal-Tanner equation in Eq (7), using numerical integration with  $n_{grid} = 10^3$  grid points to compute the values of the signal attenuation. The model is fit to the experimental NMR signal attenuation using nonlinear least-squares [20].

Bootstrapping [30] is used for obtaining confidence bounds for the conventionally estimated scaling law relationship. The eight measurement points are resampled with replacement  $n_{boot} = 10^5$  times, and 2.5% and 97.5% percentiles are computed pointwise along the  $M$  axis to yield a 95% nonparametric confidence interval.

Data analysis is performed using Matlab R2015a (Mathworks, Natick, MA, US).

## Supporting Information

**S1 Code and Data Sets. Matlab programs and data sets used in this study.**  
(ZIP)

## Acknowledgments

This research was funded by the South Australian Government Premier's Research and Industry Grant project 'A Systems Approach to Surface Science' and partly by the Australian Research Council Centre of Excellence in Convergent Bio-Nano Science and Technology (project number CE140100036). Helena Andersson is acknowledged for performing the GPC measurements.

## Author Contributions

Conceived and designed the experiments: MR SJB NHW MRD TN MN. Performed the experiments: SJB NHW MRD. Analyzed the data: MR. Contributed reagents/materials/analysis tools: SJB NHW MRD TN. Wrote the paper: MR SJB NHW MRD TN MN.

## References

1. Callaghan PT, Arns CH, Galvosas P, Hunter MW, Qiao Y, Washburn KE. Recent Fourier and Laplace perspectives for multidimensional NMR in porous media. *Magn Reson Imaging*. 2007; 25:441–444. doi: [10.1016/j.mri.2007.01.114](https://doi.org/10.1016/j.mri.2007.01.114) PMID: [17466759](https://pubmed.ncbi.nlm.nih.gov/17466759/)
2. Bernin D, Topgaard D. NMR diffusion and relaxation correlation methods: New insights in heterogeneous materials. *Curr Opin Colloid Interface Sci*. 2013; 18:166–172. doi: [10.1016/j.cocis.2013.03.007](https://doi.org/10.1016/j.cocis.2013.03.007)
3. Brixner T, Stenger J, Vaswani HM, Cho M, Blankenship RE, Fleming GR. Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature*. 2005; 434:625–628. doi: [10.1038/nature03429](https://doi.org/10.1038/nature03429) PMID: [15800619](https://pubmed.ncbi.nlm.nih.gov/15800619/)
4. Turner DB, Arpin PC, McClure SD, Ulness DJ, Scholes GD. Coherent multidimensional optical spectra measured using incoherent light. *Nat Commun*. 2013; 4. doi: [10.1038/ncomms3298](https://doi.org/10.1038/ncomms3298)
5. Ostroumov EE, Mulvaney RM, Cogdell RJ, Scholes GD. Broadband 2D electronic spectroscopy reveals a carotenoid dark state in purple bacteria. *Science*. 2013; 340:52–56. doi: [10.1126/science.1230106](https://doi.org/10.1126/science.1230106) PMID: [23559244](https://pubmed.ncbi.nlm.nih.gov/23559244/)
6. Khaleghi B, Khamis A, Karray FO, Razavi SN. Multisensor data fusion: A review of the state-of-the-art. *Inform Fusion*. 2013; 14:28–44. doi: [10.1016/j.inffus.2011.08.001](https://doi.org/10.1016/j.inffus.2011.08.001)
7. Oliva J, Póczos B, Schneider J. Distribution to distribution regression. In: *Proceedings of The 30th International Conference on Machine Learning*; 2013. p. 1049–1057.
8. Resch-Genger U, Grabolle M, Cavaliere-Jaricot S, Nitschke R, Nann T. Quantum dots versus organic dyes as fluorescent labels. *Nat Methods*. 2008; 5:763–775. doi: [10.1038/nmeth.1248](https://doi.org/10.1038/nmeth.1248) PMID: [18756197](https://pubmed.ncbi.nlm.nih.gov/18756197/)
9. Riegler J, Nick P, Kielmann U, Nann T. Visualizing the self-assembly of tubulin with luminescent nanorods. *J Nanosci Nanotechnol*. 2003; 3:380–385. doi: [10.1166/jnn.2003.219](https://doi.org/10.1166/jnn.2003.219) PMID: [14733147](https://pubmed.ncbi.nlm.nih.gov/14733147/)
10. Riegler J, Nann T. Application of luminescent nanocrystals as labels for biological molecules. *Anal Bioanal Chem*. 2004; 379:913–919. PMID: [15278340](https://pubmed.ncbi.nlm.nih.gov/15278340/)

11. Bear JC, Hollingsworth N, McNaughton PD, Mayes AG, Ward MB, Nann T, et al. Copper-Doped CdSe/ZnS Quantum Dots: Controllable Photoactivated Copper (I) Cation Storage and Release Vectors for Catalysis. *Angew Chem Int Ed*. 2014; 53:1598–1601. doi: [10.1002/anie.201308778](https://doi.org/10.1002/anie.201308778)
12. Medintz IL, Uyeda TH, Goldman ER, Mattoussi H. Quantum Dot Bioconjugates for Imaging, Labelling and Sensing. *Nat Mater*. 2005; 4:435–446. doi: [10.1038/nmat1390](https://doi.org/10.1038/nmat1390) PMID: [15928695](https://pubmed.ncbi.nlm.nih.gov/15928695/)
13. Shen J, Zhu Y, Yang X, Li C. Graphene quantum dots: Emergent nanolights for bioimaging, sensors, catalysis and photovoltaic devices. *Chem Comm*. 2012; 48:3686–3699. doi: [10.1039/c2cc00110a](https://doi.org/10.1039/c2cc00110a) PMID: [22410424](https://pubmed.ncbi.nlm.nih.gov/22410424/)
14. Kumar S, Nann T. First solar cells based on CdTe nanoparticle/MEH-PPV composites. *J Mater Res*. 2004; 19:1990–1994. doi: [10.1557/JMR.2004.0279](https://doi.org/10.1557/JMR.2004.0279)
15. Lakowicz JR. *Principles of Fluorescence Spectroscopy*. Springer; 2006.
16. de Gennes PG. *Scaling concepts in polymer physics*. Cornell University Press; 1979.
17. Flory PJ. *Principles of polymer chemistry*. Cornell University Press; 1953.
18. Doi M, Edwards SF. *The theory of polymer dynamics*. Oxford University Press; 1988.
19. Augé S, Schmit PO, Crutchfield CA, Islam MT, Harris DJ, Durand E, et al. NMR measure of translational diffusion and fractal dimension. Application to molecular mass measurement. *The Journal of Physical Chemistry B*. 2009; 113:1914–1918. doi: [10.1021/jp8094424](https://doi.org/10.1021/jp8094424) PMID: [19173563](https://pubmed.ncbi.nlm.nih.gov/19173563/)
20. Williamson NH, Nydén M, Röding M. The lognormal and gamma distribution models for estimating molecular weight distributions of polymers using PGSE NMR. Accepted for publication in *Journal of Magnetic Resonance*. 2016;. doi: [10.1016/j.jmr.2016.04.007](https://doi.org/10.1016/j.jmr.2016.04.007)
21. Håkansson B, Nydén M, Söderman O. The influence of polymer molecular-weight distributions on pulsed field gradient nuclear magnetic resonance self-diffusion experiments. *Colloid and Polymer Science*. 2000; 278:399–405. doi: [10.1007/s003960050532](https://doi.org/10.1007/s003960050532)
22. Stejskal EO, Tanner JE. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*. 1962; 42:288–298. doi: [10.1063/1.1695690](https://doi.org/10.1063/1.1695690)
23. Gong X, Hansen EW, Chen Q. A simple access to the (log/normal) molecular weight distribution parameters of polymers using PGSTE NMR. *Macromolecular Chemistry and Physics*. 2011; 212:1007–1015. doi: [10.1002/macp.201000706](https://doi.org/10.1002/macp.201000706)
24. Ziegler J, Merkulov A, Grabolle M, Resch-Genger U, Nann T. High-quality ZnS shells for CdSe nanoparticles: Rapid microwave synthesis. *Langmuir*. 2007; 23:7751–7759. doi: [10.1021/la063528b](https://doi.org/10.1021/la063528b) PMID: [17552544](https://pubmed.ncbi.nlm.nih.gov/17552544/)
25. Callaghan PT, Pinder DN. Influence of polydispersity on polymer self-diffusion measurements by pulsed field gradient nuclear magnetic resonance. *Macromolecules*. 1985; 18:373–379. doi: [10.1021/ma00145a013](https://doi.org/10.1021/ma00145a013)
26. Callaghan PT. *Translational dynamics and magnetic resonance: principles of pulsed gradient spin echo NMR*. Oxford University Press; 2011.
27. Pawitan Y. *In all Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford University Press; 2001.
28. Röding M, Bradley SJ, Nydén M, Nann T. Fluorescence Lifetime Analysis of Graphene Quantum Dots. *The Journal of Physical Chemistry C*. 2014; 118:30282–30290. doi: [10.1021/jp510436r](https://doi.org/10.1021/jp510436r)
29. Robert CP, Casella G. *Monte Carlo Statistical Methods*. Springer; 2013.
30. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. CRC press; 1994.