Characterization of colloidal nanocrystal surface structure using small angle neutron scattering and efficient Bayesian parameter estimation

Cite as: J. Chem. Phys. **150**, 244702 (2019); https://doi.org/10.1063/1.5108904
Submitted: 03 May 2019 . Accepted: 04 June 2019 . Published Online: 24 June 2019

Samuel W. Winslow, Wenbi Shcherbakov-Wu, Yun Liu, William A. Tisdale, and James W. Swan

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Characterization of colloidal nanocrystal surface structure using small angle neutron scattering and efficient Bayesian parameter estimation

Cite as: J. Chem. Phys. 150, 244702 (2019); doi: 10.1063/1.5108904
Submitted: 3 May 2019 • Accepted: 4 June 2019 •
Published Online: 24 June 2019

Samuel W. Winslow,[1] Wenbi Shcherbakov-Wu,[2] Yun Liu,[3,a)] William A. Tisdale,[1,b)] and James W. Swan[1,c)]

AFFILIATIONS

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA
[2]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA
[3]Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

Note: This paper is part of the JCP Special Topic on Colloidal Quantum Dots.
[a)]**Also at:** Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, USA.
[b)]**Electronic mail:** tisdale@mit.edu
[c)]**Electronic mail:** jswan@mit.edu

ABSTRACT

Complete structural characterization of colloidal nanocrystals is challenging due to rapid variation in the electronic, vibrational, and elemental properties across the nanocrystal surface. While electron microscopy and X-ray scattering techniques can provide detailed information about the inorganic nanocrystal core, these techniques provide little information about the molecular ligands coating the nanocrystal surface. Moreover, because most models for scattering data are parametrically nonlinear, uncertainty estimates for parameters are challenging to formulate robustly. Here, using oleate-capped PbS quantum dots as a model system, we demonstrate the capability of small angle neutron scattering (SANS) in resolving core, ligand-shell, and solvent structure for well-dispersed nanocrystals using a single technique. SANS scattering data collected at eight separate solvent deuteration fractions were used to characterize the structure of the nanocrystals in reciprocal space. Molecular dynamics simulations were used to develop a coarse-grained form factor describing the scattering length density profile of ligand-stabilized nanocrystals in solution. We introduce an affine invariant Markov chain Monte Carlo method to efficiently perform nonlinear parameter estimation for the form factor describing such dilute solutions. This technique yields robust uncertainty estimates. This experimental design is broadly applicable across colloidal nanocrystal material systems including emergent perovskite nanocrystals, and the parameter estimation protocol significantly accelerates characterization and provides new insights into the atomic and molecular structure of colloidal nanomaterials.

I. INTRODUCTION

Characterization techniques for colloidal semiconducting quantum dots (QDs) such as PbS, CdSe, or InP are typically sensitive to properties of the core. Common core-characterization techniques such as small angle X-ray scattering (SAXS), transmission electron microscopy (TEM), and UV-Vis spectroscopy[1–3] are not sensitive to the organic capping ligands, which are vital to impart colloidal stability[4] and engineered to tune inter-QD spacing in devices with QD thin films.[5] Although the QD core composition principally determines the optoelectronic properties,[6–11] recent work has highlighted the importance of the surface-bound ligands in determining the structural, optical, and electronic properties of QDs.[12,13] For instance, the core surface has been shown to have a strong effect on electron-phonon coupling.[14,15] Additionally, ligand choice and surface coverage control superlattice morphology in thin films.[5,16–24] However, to elucidate the manner in which ligand coverage affects superlattice structure, the ligand coverage must be known. One technique sensitive to the ligand environment is nuclear magnetic resonance (NMR) spectroscopy. Quantitative NMR may measure

a ligand coverage, but this data interpretation must be supplemented with other techniques sensitive to the core material.[25–27] A single measurement capable of fully quantifying colloidal QDs' structure remains absent from the experimental arsenal available to researchers.

Small angle neutron scattering (SANS) can resolve organic ligands separately from the inorganic core and surrounding solvent due to the controllable contrast between hydrogen-bearing materials and the solvent molecules. The scattering length density (SLD) of a material, which is the total coherent scattering length of a molecule normalized by its molecular volume, determines the observed intensity profile. Contrast or resolution stems from the difference between the SLDs of each material in a nanocrystal and its ligand-shell and the solvent. Moreover, the ability to tune solvent contrast with mixtures of deuterated solvent allows SANS experiments to resolve specific components. While model-independent analysis yields descriptors of the probed system such as fractal dimension or specific surface area, precise descriptions of material structure can only be determined after fitting experimental data to a parameterized form factor over the full range of scattering vectors.

Typical nonlinear parameter estimation algorithms attempt to minimize the sum of square residuals between the data and the model predictions. Algorithms such as the Levenberg-Marquardt method, which deploys trust-region methods, can be used to converge quickly to a minimum.[28,29] These least-squares methods attempt to minimize the sum of squared residuals between the data and model prediction. The solutions found may be only local solutions, and uncertainty estimates are based on a quadratic approximation of the objective function at the minimum. This approximation of the local solution landscape may not reflect the experimental uncertainty. Bayesian inference is a more principled way of parameter estimation by formulating a statistical distribution for the parameters inherited from the experimental uncertainty. Specifically, Markov chain Monte Carlo (MCMC) methods may be used to sample from the underlying parameter distributions and to measure quantities such as the mean and most likely parameter values as well as generate samples of the parameters that are representative of their uncertainty with no *ad hoc* approximations required.

In the cosmology literature, MCMC methods have been used successfully for parameter estimation of cosmic microwave background radiation of at least 10 parameters,[30] with convergence of the Markov chain verified by spectral methods.[31] A particular MCMC method, which leverages affine invariance to efficiently take steps with multiple chains in parallel as opposed to one chain, was introduced in 2010 by Goodman and Weare[32] and implemented in a Python package *emcee* by Foreman-Mackey *et al.* in 2012.[33] This new implementation has already been deployed for increased efficiency of cosmological parameter estimation.[34] We endeavor in this work to similarly provide a more efficient data fitting method for the neutron scattering community.

Though Bayesian inference methods are already available in some SANS data fitting packages, their implementation is not efficient or flexible enough for complicated form factors. The *DREAM* algorithm available for SANS data fitting in the *SasView* software adaptively reinitializes initial parameter samples to increase likelihood of finding a global minimum.[35] However, when a custom form factor is required, as is posed in this work, there is a high barrier to efficient implementation. Another Monte Carlo based small angle scattering software package *McSAS* is also available for use with these material systems.[3,36] However, the method is slow for large numbers of parameters and does not permit incorporation of custom form factors. The affine invariant MCMC method of Goodman and Weare functions well for the purposes of custom form factors with large numbers of parameters. The obtained parameter distributions, or posteriors, allow propagation of experimentally derived uncertainty in measured quantities derived from a model fit.

We apply the Goodman and Weare method to neutron scattering. The form factor proposed in this study is directly inspired by the relaxed structure of PbS QDs sampled from molecular dynamics (MD) simulations. This core material is selected as a prototypical system useful to demonstrate the capability of the new method. In this study, we fully quantify the QDs, including the QD surface, with a single SANS solvent variation experiment. Data fitting with the affine invariant MCMC method yields realistic estimates of parameters characterizing the form factor and their uncertainties. The fit QD structure corresponds well to MD simulation, justifying the proposed form factor. The data and executable MATLAB code used to perform the MCMC fit are included in the supplementary material.

II. MATERIALS AND METHODS

To fit SANS data, a form factor describing particle shape is needed. The form factor is used to calculate the expected scattering intensity profile. In this study, existing form factors available in fitting software libraries were found insufficient. As such, a custom form factor was developed. To ensure the proposed form factor could capture any rational variation in a batch of QDs, the analytical form was based on single-QD MD simulations in which shape, size, and ligand coverage could be manipulated. Here, we describe the details for MD simulation; QD synthesis and sample preparation; as well as SANS data collection, treatment, and fitting.

A. MD simulation

PbS is a semiconducting material with the NaCl crystal lattice in the bulk. As a QD, the material does not have the bulk cubic shape but is polyhedral with exposed {100} and {111} crystal planes. PbS model cores were constructed by cutting away the corners of a rock-salt structure bulk crystal along the {111} planes to leave a specified number of Pb atoms along the edges of adjoining (111)/(111), (111)/(100), and (100)/(100) facets. This allowed any size core in the full truncation sequence of an octahedron, truncated octahedron, cuboctahedron, truncated cube, and cube to be generated. (111) facets are completely Pb-terminated, while (100) facets are stoichiometric. Ligands were grafted randomly onto (111) Pb atoms at a specified coverage, in line with previous computational[18] and experimental[37] findings that ligands bind preferentially to the (111) core facets. Precise binding ratios are not known, but in this study we affix ligands solely to the (111) facets. Charge is not included in the simulations. The united atom (UA) representation was used to coarse grain CH_x groups for ligands on the surface[38] and toluene

solvent.[39] Parameters describing interactions of the core with ligands or solvent were modeled after simulations of PbSe QDs.[40–42] The interaction energy and distance for lead were identical to those works, while the parameters for sulfur were back-calculated using Lorentz-Berthelot mixing rules from simulations of CdSe and CdS QDs.[43,44] Simulations were run in the NVT ensemble using the HOOMD-blue software package[45–47] using mass, length, and energy scales of 1 u, 1 Å, and 1 kJ/mol, respectively. With these units, the simulation inherent time scale was 0.1 ns. A time step of 5 fs was used. Toluene was initialized on a simple cubic lattice at its room temperature density. A single QD was added to the middle of the simulation box. Overlapping toluene molecules were removed. The simulation box was large enough so that the QD did not interact with itself beyond the 10 Å cutoff distance. Each simulation was equilibrated for 100 ns before a production run of 100 ns. Following simulation, the bead distances from the QD center of mass were calculated and averaged. Profiles of distance from the QD center of mass for each material were constructed from histograms of bin width 4 Å.

B. Synthesis, characterization, and sample preparation

PbS QDs were synthesized using a modification of the hot-injection protocols from Hines and Scholes[48] and Colbert *et al.*[49] 450 mg (2.0 mmol) of PbO (Sigma-Aldrich, ≥99%), 14 g of 1-octadecene (Sigma-Aldrich, 90%), and 1.4 g (4.96 mmol) of oleic acid (Sigma-Aldrich, 90%) were used for the lead precursor. The mixture was stirred under vacuum at 120 °C for 3 h, turning the solution clear. The flask was placed under flowing nitrogen gas, and the temperature was raised to 130 °C. Meanwhile, the sulfur precursor was prepared in a second three-neck flask with 4 g of 1-octadecene and 210 μl of hexamethyldisilathiane (HMDS) (Sigma-Aldrich, synthesis grade). Following injection, the heating mantle was removed after 5 s, and the flask was cooled by a room-temperature water bath. The QDs were isolated by precipitation with acetone followed by centrifugation. The product was then cleaned in ambient air by dissolving the precipitated QDs in ~3 ml of hexanes and precipitating once with acetone and twice with methanol. The precipitated QDs were then brought into the glovebox and re-dispersed in anhydrous toluene.

QD size and solution concentration were calculated from UV-Vis absorbance measurements.[2,3] The measured size was 3.41 nm ($\sigma_d \leq 4.5\%$).[1] ^{1}H-NMR was performed on a three-channel Bruker Avance Neo spectrometer operating at 500.34 MHz equipped with a 5 mm liquid-nitrogen cooled Prodigy broadband observe (BBO) cryoprobe. The instrument runs in full-automation off of a SampleXpress 60 sample charger. Known quantities of QDs, measured by absorbance,[1] were dissolved in 650 μl of deuterated toluene containing a known quantity (2 μl) of dibromomethane as an internal quantitative standard. Measurement temperature was 25 °C. The spectra were measured under conditions of full T_1 relaxation (delay of 45 s). The integrated area under the CH_2Br_2 resonance and the ligand vinyl proton resonance were used to measure the number of ligands per QD.[50] Figure S1 shows the solvent, standard, and ligand peaks in a typical spectrum.

SANS samples were prepared by evaporating off the solvent from an aliquot of known concentration overnight in a vacuum chamber before redispersion in a measured mixture of hydrogen-bearing (ACS, ≥99.5%) and deuterated toluene (Sigma-Aldrich, ≥99.6% C_7D_8) to an approximate total scatterer volume fraction of 1.0 v/v% so that the solution may be considered dilute. The concentration is the same for each sample. A ligand coverage of 3 ligands/nm^2 was assumed for this calculation. It is only necessary to estimate the volume fraction as the true volume fraction is fit from the data. Eight different deuteration fractions were used ranging from 20% to 100% in 10% intervals, excluding 30% due to beamtime limits.

C. SANS instrument and data reduction

SANS experiments were performed on the NGB30mSANS instrument at the NIST Center for Neutron Research (NCNR).[51] Three detector distances were used to cover the q range from approximately 0.0055 Å^{-1} to about 0.445 Å^{-1}. The neutron wavelength for all detector configurations was 6 Å with a wavelength spread of 12%. The samples were loaded into demountable titanium cells with quartz windows provided by NCNR. The obtained data were reduced using the SANS data reduction package of NCNR.[52]

D. Analysis of SANS data

1. Intensity profile

The observed intensity profile $I(\mathbf{q})$ in a SANS experiment for dilute, monodisperse solutions may be written as

$$I(\mathbf{q}) = \frac{\phi}{V_p} F(\mathbf{q}) F(\mathbf{q})^* = \frac{\phi}{V_p} |F(\mathbf{q})|^2. \tag{1}$$

Here, ϕ is the volume fraction of all scatterers, V_p is the volume of the scattering particle, and $F(\mathbf{q})$ is termed the shape factor describing the geometry of the particle. For a noncentrosymmetric object, $F(\mathbf{q})$ may have an imaginary part, but for a centrosymmetric scatterer $F(\mathbf{q})$ may be assumed real. The commonly referred to form factor is $P(\mathbf{q}) = |F(\mathbf{q})|^2$. An incoherent background term is absent from this equation but would enter as an added q-independent constant term. The shape factor is written as the Fourier transform of the excess SLD profile $\Delta\rho(\mathbf{r})$,

$$F(\mathbf{q}) = \mathscr{F}(\Delta\rho(\mathbf{r})) = \int_V \Delta\rho(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r}. \tag{2}$$

The vector $\mathbf{r}$ covers 3D space. We change to spherical coordinates and integrate the angular dependences to arrive at

$$F(q) = 4\pi \int_0^\infty \Delta\rho(r) r^2 j_0(qr) dr. \tag{3}$$

Here, j_0 denotes the zeroth order spherical Bessel function of the first kind. The excess SLD is fixed for any distance r measured from the QD core center. For any model choice of $\Delta\rho(r)$, the integration can be performed numerically or analytically to find the intensity profile $I(q)$,

$$I(q) = \frac{\phi}{V_p} \left(4\pi \int_0^\infty \Delta\rho(r) r^2 j_0(qr) dr \right)^2. \tag{4}$$

This treatment approximates the anisotropic QD as an isotropic object such that orientation does not matter. This is done because of the high sphericity of QDs in the size range of interest and the ensemble nature of the technique yielding an intensity profile

already averaged over QD orientation. The choice of model SLD profile is described in Sec. III A. We use a piecewise linear $\Delta\rho(r)$ profile for which each linear section can be directly integrated. This approximation requires a set of points $\{r_k\}$ and their corresponding SLDs $\{\Delta\rho(r_k)\}$. The total intensity is a sum of the integral in each linear section. A linear profile connecting the points (r_k, ρ_k) and (r_{k+1}, ρ_{k+1}) with the form

$$\Delta\rho_{lin}(r) = (\rho_k - \rho_{solv}) + \left(\frac{\rho_{k+1} - \rho_k}{r_{k+1} - r_k}\right)(r - r_k) \tag{5}$$

will yield the scattering intensity

$$I(q) = \frac{\phi}{V_p}\left(\sum_{k=1}^{N-1} 4\pi \int_{r_k}^{r_{k+1}} \Delta\rho_{lin}(r) r^2 j_0(qr) dr\right)^2. \tag{6}$$

Because the integration may be done analytically, data fitting is accelerated and accuracy of the model is high. The derivation of and full expression for $I(q)$ when $\Delta\rho(r)$ is a piecewise linear function are given in Appendix A.

2. Polydispersity and instrument smearing

Convolution integrals describing parameter polydispersity take the form

$$\hat{I}(q) = \int_0^\infty D(q;\theta) I(q) d\theta, \tag{7}$$

where $\hat{I}(q)$ denotes a transformed intensity profile and $D(q;\theta)$ denotes an assumed distribution in some parameter θ. Typical distributions in scattering experiments are Gaussian or Schulz. The integral sums each contribution to the observed intensity weighted by the value of the assumed distribution at each point. These integrals are rarely analytically computable and are expensive to compute numerically, requiring recalculation at each iteration of a fitting algorithm. Such a convolution is useful when performing nonlinear least squares fitting but unnecessary for Bayesian methods of parameter estimation. For the Bayesian approach used here, distributions of each parameter are obtained inherently by the MCMC data fitting method sampling from the posterior. The supplementary material discusses when omission of polydispersity convolution integrals is acceptable. Section III C discusses the parameter distributions and their analogy to polydispersity.

Accounting for instrument smearing is necessary due to the finite size of detector pixels and pinhole collimation of the neutron beam. The intensity from the 2D detector is azimuthally integrated. We denote the list of wave vector magnitudes at which the intensity was sampled $\mathbf{q}$. For each sampled q_i, a Gaussian function describing the resolution is known through calibration and is parameterized by a mean $\bar{q}_i$ and standard deviation σ_i. The observed, smeared intensity profile $I_{smear}(q_i)$ at each q_i is given by

$$I_{smear}(q_i) = \int_{-\infty}^{\infty} R(q';\bar{q}_i,\sigma_i) I(q') dq'. \tag{8}$$

$R(q';\bar{q}_i,\sigma_i)$ is the Gaussian resolution function for each q_i,

$$R(q';\bar{q},\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(q'-\bar{q})^2}{2\sigma^2}\right). \tag{9}$$

This integral is of a form suitable for efficient numerical evaluation by Gauss-Hermite quadrature. The analytical form of the intensity profile was smeared with this implementation to match the experiment. This process is described in Appendix B.[53]

E. MCMC implementation

We estimate the model parameters using a Bayesian point of view. To infer a set of model parameters $\boldsymbol{\theta}$ after collecting data $\mathbf{I}$, a vector of scattering intensities, we utilize Bayes' theorem,

$$p(\boldsymbol{\theta}|\mathbf{I}) \propto p(\mathbf{I}|\boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{10}$$

where $p(\boldsymbol{\theta}|\mathbf{I})$ is the posterior probability of the parameters given the measured data, $p(\mathbf{I}|\boldsymbol{\theta})$ is the probability of observing the data given a set of parameters, and $p(\boldsymbol{\theta})$ represents prior knowledge of the parameters $\boldsymbol{\theta}$. When $p(\mathbf{I}|\boldsymbol{\theta})$ is viewed as a function of $\mathbf{I}$ with the true parameters fixed, it is a probability density function. From the perspective where $p(\mathbf{I}|\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ with the data having been collected, it is a likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{I})$. The likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{I})$ informs the posterior probability of the parameters $p(\boldsymbol{\theta}|\mathbf{I})$ and is modulated by any previous knowledge of the parameters contained in $p(\boldsymbol{\theta})$.

We assume that the data, collected from the mean scattering intensity, are normally distributed about the model. Then, for each measurement $I(q_i)$,

$$p(I(q_i)|\boldsymbol{\theta}) \propto \exp\left(-\frac{\left(I(q_i)^{\text{exp}} - I(q_i;\boldsymbol{\theta})^{\text{model}}\right)^2}{\Sigma_{I,i}}\right), \tag{11}$$

where $I(q_i)^{\text{exp}}$ is the measured intensity, $I(q_i;\boldsymbol{\theta})^{\text{model}}$ is calculated from the smeared intensity profile previously described, and $\Sigma_{I,i}$ represents the variance of the data at q_i. The joint probability of observing the full set of data $\mathbf{I}$ given parameters $\boldsymbol{\theta}$ is the product of the individual $p(I(q_i)|\boldsymbol{\theta})$ over all data points,

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{I}) = p(\mathbf{I}|\boldsymbol{\theta}) \propto \prod_{i=1}^{N} p(I(q_i)|\boldsymbol{\theta}). \tag{12}$$

This specifies the likelihood a set of parameters describes the experiment given the data. By taking the logarithm of both sides, we find that the log-likelihood may be given by the negative weighted sum of squared error,

$$\ln(\mathcal{L}(\boldsymbol{\theta}|\mathbf{I})) \propto -\sum_{i=1}^{N}\left(\frac{I(q_i)^{\text{exp}} - I(q_i;\boldsymbol{\theta})^{\text{model}}}{\Sigma_{I,i}^{1/2}}\right)^2. \tag{13}$$

In general, an MCMC algorithm samples from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{I})$ by constructing a chain of positions in parameter space with each subsequent position $\boldsymbol{\theta}_{t+1}$ being dependent solely on the previous step $\boldsymbol{\theta}_t$. With steps taken appropriately, the distribution of the steps in this Markov chain will represent the underlying parameter distribution, given the observed measurements.

We implement the MCMC method introduced by Goodman and Weare,[32] which leverages the property of affine invariance to improve performance. We utilize the MATLAB algorithm *gwmcmc* of Grinsted[54] which is similar to the Python implementation, *emcee*, developed by Foreman-Mackey *et al.*[33] This implementation exploits the ensemble nature of the method with parallel

calculation over multiple Markov chains simultaneously. This study uses 100 chains.

In each step, the chains are partitioned into two sets $S^{(0)}$ and $S^{(1)}$. A proposed state $\boldsymbol{\theta}_{t+1}^{(0)\prime}$ for a walker in set $S^{(0)}$ is selected randomly along a ray connecting its current position $\boldsymbol{\theta}_t^{(0)}$ with the position of a randomly selected walker of set $S^{(1)}$,

$$\boldsymbol{\theta}_{t+1}^{(0)\prime} = \boldsymbol{\theta}_t^{(0)} + Z\left[\boldsymbol{\theta}_t^{(1)} - \boldsymbol{\theta}_t^{(0)}\right]. \tag{14}$$

The random variable Z is drawn from a probability distribution $g(Z = z)$. If $g(z)$ satisfies $g(z^{-1}) = zg(z)$, then the proposal is symmetric. Goodman and Weare[32] advocated for a particular form of $g(z)$,

$$g(z) \propto \begin{cases} \dfrac{1}{\sqrt{z}}, & \text{if } z \in \left[\dfrac{1}{a}, a\right], \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

where a is a tunable step size set to 1.25 in this study. To satisfy the detailed balance condition, the proposed state is accepted according to

$$\boldsymbol{\theta}_{t+1}^{(0)} \leftarrow \begin{cases} \boldsymbol{\theta}_{t+1}^{(0)\prime}, & \text{if } z^{d-1}\dfrac{\mathcal{L}(\boldsymbol{\theta}_{t+1}^{(0)\prime}|\mathbf{I})p(\boldsymbol{\theta}_{t+1}^{(0)\prime})}{\mathcal{L}(\boldsymbol{\theta}_t^{(0)}|\mathbf{I})p(\boldsymbol{\theta}_t^{(0)})} \geq r, \\ \boldsymbol{\theta}_t^{(0)}, & \text{otherwise.} \end{cases} \tag{16}$$

Here, r is a random number from 0 to 1 and d is the number of parameters in the model. The step size a is chosen to obtain an acceptance fraction between 25% and 50%, indicating the walkers are efficiently traversing parameter space. It has been found that the optimal acceptance rate for efficient sampling tends to 25% for models of four or more parameters.[31] In this work, we assume an uninformative prior such that any parameter value within prescribed bounds is equally likely.

The performance of this particular proposal scheme is independent of the underlying posterior distribution.[32] With enough samples, it is guaranteed to recover the posterior. In practice, the chains are run for a finite number of steps. Statistics must only be computed on a stationary distribution when the chains are only sampling the posterior locally. If not, then the computed statistics will not represent the parameter distributions as predicted by the data. Foreman-Mackey *et al.* suggested running the simulation for at least 10 autocorrelation times to obtain independent samples from the posterior.[33] The autocorrelation times for this system across all batches range from 4780 to 6340 iterations. Following Foreman-Mackey *et al.*, at least 63 400 total samples are needed. Using 100 Markov chains, it would be sufficient to run the simulation for at least 634 iterations.

In generating production data, starting the chains in a tight region around a solution would allow rapid convergence to the posterior. Without knowing this solution *a priori*, a preliminary run is first performed with walker initial positions sampled randomly over an uninformative prior. The portion of the chain where walkers have not yet converged is discarded as a "burn-in" period because the distribution is not yet stationary. The first 1000 iterations per chain are discarded. The next 2000 iterations serve to sample from the posterior. It is possible that walkers sample local solutions that do not predict the data by manual inspection. These walkers are discarded, and a run is re-initialized in a tight ball around the remaining walkers within 5% of the current walker with the highest likelihood.[33] A subsequent 3000 iterations, more than enough to produce independent samples, represent the production data. Additional discussion of practical considerations may be found in the supplementary material.

III. RESULTS AND DISCUSSION

A. Proposed form factor from MD simulation

MD simulation was used to obtain the relaxed structure of an isolated QD in solution as a model for the shape factor $F(q)$ for fitting SANS data. Following equilibration of the simulation, the angularly integrated radial volume fraction profile $\varphi(r)$ of each material (PbS, ligand, solvent) was calculated from histograms of coarse-grained bead positions. This profile is equivalent to a number density profile but is convenient in introducing constraints on the sum of volume fractions. Bulk PbS has a cubic shape with a Pb:S ratio of 1:1. However, as the QDs nucleate and grow, competition between adding atoms to (100) and (111) facets yields polyhedral core shapes. The flexibility of the MD model allows initialization of any core shape in the growth sequence of PbS from an octahedron, truncated octahedron, cuboctahedron, and ultimately cube. Expected core shapes as the QDs grow are illustrated in Fig. 1. For any Pb:S ratio and total number of atoms, there is a uniquely shaped core. In the size range of 3–8 nm typical for PbS QDs, the truncated octahedron is expected to represent the core well. In this size range, Pb:S ratios are typically 1.1–1.3.[50,55] The exact core shape is not known. In order to develop a flexible form factor model capable of describing any batch of QDs, core sizes and shapes spanning the expected growth path were simulated and ligand coverage on the core surface was varied. Figure 2(a) shows the volume fraction profiles for one truncated octahedron shaped QD with three different ligand coverage specifications. From the perspective of the radial volume fraction profile, there is an overlap of core material, ligand, and solvent in the vicinity of the QD

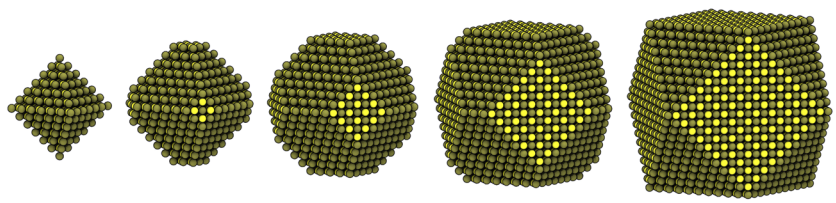

FIG. 1. Possible growth of PbS QD cores from an approximate 2.5 nm octahedron through a 4.7 nm truncated octahedron to a 6.9 nm cuboctahedron. Sizes are based on equivalent spherical diameter. Exposed (111) Pb-terminated facets reduce in fraction of total surface area in favor of stoichiometric (100) facets.

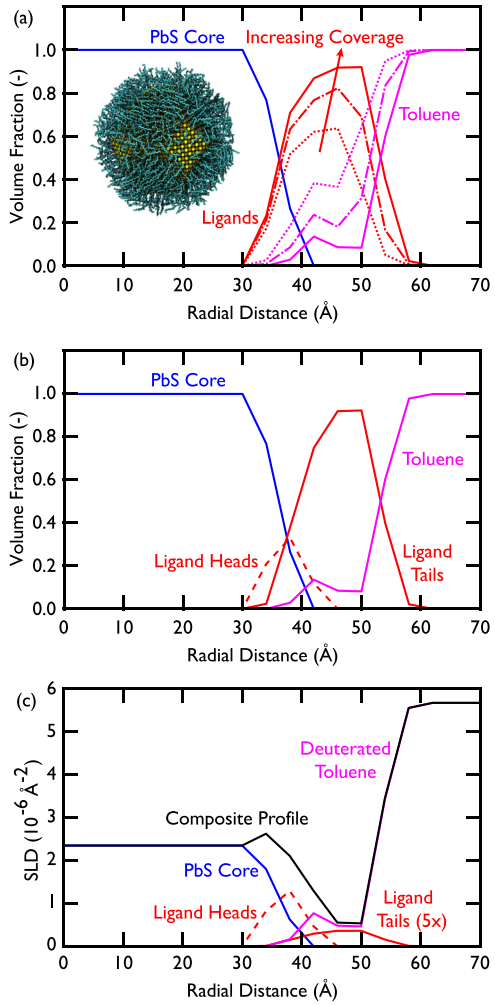

FIG. 2. Extraction of volume fraction and SLD profiles from MD simulation of a 7 nm truncated octahedral QD. (a) Volume fraction profiles with 3 separate ligand coverages. Bin widths are 4 Å. As ligand coverage increases, solvent is excluded from the shell. The inset shows the core and ligands' MD structure from the middle ligand coverage of approximately 4.2 ligands/nm^2. (b) Volume fraction profile treating the ligand head (red, dashed) and tail (red, solid) groups separately. (c) Corresponding SLD profile using deuterated toluene. Contributions to the total profile are shown for the PbS core (blue), ligand tails (red, solid), ligand heads (red, dashed), and solvent (magenta). The ligand tail contribution is multiplied by 5 to show its shape. Almost all ligand contribution to the SLD profile at the surface comes from the head groups.

surface due to the faceted shape of the QD and spherical integration. This is the first indication that a simple spherical description of the QDs with sharp interfaces between materials as in a core-shell model will not be sufficient. As ligand coverage is increased, solvent is excluded from the ligand shell. However, a population of solvent does remain near to the surface. Preferential attachment of ligands to the (111) facets leaves void space over the (100) facets where solvent can reside.[18,37] Additionally, some solvent will penetrate the ligands to solvate the QD. At higher ligand coverage, the transition between ligand and solvent phases becomes sharper. Steric hindrance forces solvent out of the ligand shell.

It is necessary to treat the ligand head group bound to the nanocrystal surface separately from the solvated tails. The ligands are bound as oleate molecules, with a CO_2^- group binding at surface lead sites; oxygen is a more efficient neutron scatterer than carbon or hydrogen in that it has a higher SLD. To probe the potential shape of the ligand head group $\varphi(r)$ profile, the two closest ligand beads in each molecule may be integrated separately as if they were a different material. The actual distribution of the head groups is unknown, so this treatment is done to visualize the shape of a possible $\varphi(r)$ profile. Figure 2(b) shows the new $\varphi(r)$ profiles with the ligand head groups plotted separately from the tail groups. We find that different core shapes possess similar profiles. See the supplementary material for different core shape volume fraction profiles.

To calculate an expected intensity profile from scattering, the SLD profile $\varrho(r)$ is needed. The SLD is a unique property for each material and is proportional to density. As such, the SLD profile is proportional to the volume fraction profile. We convert the volume fraction profiles to SLD profiles by multiplying by the bulk SLD of each material, ϱ_i^{bulk},

$$\varrho_i(r) = \varrho_i^{\text{bulk}} \varphi(r). \tag{17}$$

Under the assumption of ideal mixing, the observed excess SLD profile $\Delta\varrho(r)$ at any radial distance will be a weighted sum of the components in the system at that distance minus the solvent SLD,

$$\Delta\varrho(r) = \left(\sum_{i=1}^{N_{comp}} \varrho_i^{\text{bulk}} \varphi_i(r) \right) - \varrho_{solvent}. \tag{18}$$

Figure 2(c) shows the expected SLD profile for the QD simulated in Fig. 2(a) with fully deuterated toluene. The SLD of dry ice at a density of 1.56 g/cm^3 is used to scale the ligand head group $\varphi(r)$ profile corresponding to the upper limit on CO_2^- head group packing density. Using less deuterated toluene would decrease contrast with the solvent and shift the SLD profile slightly where solvent penetrates the ligand shell. The profile is reminiscent of a core-shell profile. However, a core-shell profile would have sharp steps at material interfaces without capturing the material overlap. This overlap of the solvent SLD profile with that of the QD core and shell is critical to successfully predicting the data. The same parameters describing the QD must predict the data across all solvent deuteration fractions. The SLD profile calculated from MD simulation demonstrates the smooth transitions in SLD between the core, shell, and solvent. Ligand tails do not contribute much directly to the SLD profile but exclude solvent from the shell. Additionally, the ligand head groups introduce a bump up in the SLD profile at the surface. High ligand coverage will increase this bump at the core surface as well as exclude more solvent from the shell.

The MD results are used to craft a flexible model capable of capturing the volume fraction profiles of any core shape or ligand coverage. This model is used to calculate the SLD profile which is then Fourier transformed, squared, and smeared to obtain the model SANS intensity profile for data fitting. Figure 3 shows the proposed volume fraction profile and parameters to be fit to the data. Up to a distance r_{min}, it is assumed that there is only PbS. The PbS core will have a volume fraction of 0 once the distance of a vertex is reached

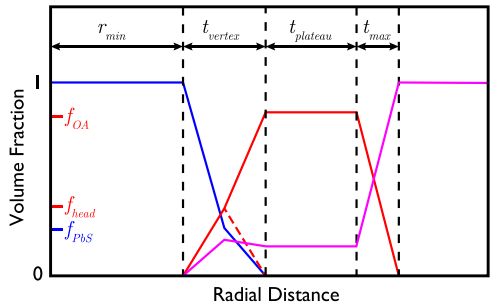

FIG. 3. Schematic of volume fractions to be fit. Only PbS (blue) is present out to a distance r_{min} before decaying over a distance t_{vertex} to 0. Near the core surface, there are contributions from PbS (f_{PbS}, blue), ligand heads (f_{head}, red), and solvent (magenta). The value of f_{head} is calculated such that the number of head groups is equal to the number of full ligand molecules. Without the ability to resolve the location of the peak, half of the PbS decay distance is used to discretize the surface layer. The ligand profile is assumed to be constant at a value f_{OA} for a distance $t_{plateau}$ in the shell before decaying linearly over a distance t_{max}. Solvent values are calculated by summing volume fractions to unity.

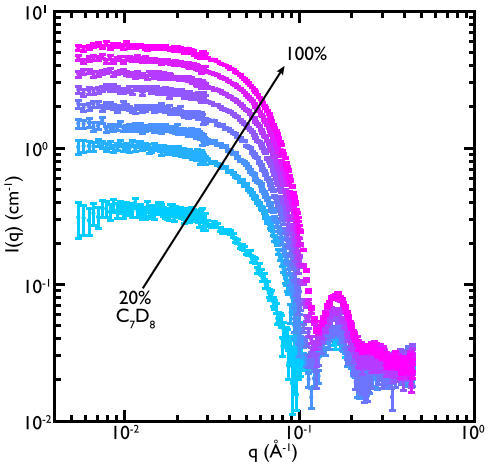

FIG. 4. Full SANS data set for nominally 3.41 nm PbS QDs in deuterated toluene. Deuteration fractions vary from 20% to 100%. As more deuterated toluene is used, the signal-to-noise ratio increases. All incoherent background from the solvent has been subtracted off, but QD incoherent background remains. Error bars represent standard deviations in intensity at each value of q.

$r_{min} + t_{vertex}$. Near the core surface, there may be scattering contributions from PbS, ligand heads, and solvent. The surface layer is discretized at a distance $r_{min} + \frac{1}{2}t_{vertex}$ to fit parameters for the volume fractions of PbS and ligand heads. In the ligand shell, the volume fraction will be at a constant value f_{OA} over a distance $t_{plateau}$. The ligand shell then decays to 0 over a distance t_{max}. Each volume fraction profile may be integrated to calculate a number of molecules, further described in Sec. III C. The value of f_{head} is calculated by requiring the number of ligand head groups to equal the number of full ligand molecules. The solvent volume fraction is calculated by summing individual volume fractions to unity. This flexible model encompasses others such as a core-shell model. Solvent penetration was elucidated by varying deuteration fraction across measurements. We scale the core, ligand tail, and solvent volume fraction profiles by the bulk SLD of each material. The SLD of dry ice is used to scale the ligand head CO_2^- group. There are then 8 model parameters: 4 distances r_{min}, t_{vertex}, $t_{plateau}$, and t_{max}; 2 volume fractions f_{PbS} and f_{OA}; 1 total volume fraction of scattering particles ϕ; and 1 constant incoherent scattering background term I_{incoh}. This model and treatment enables quantification of the core size and ligand coverage simultaneously.

B. SANS fitting with MCMC

SANS data were collected at 8 deuteration fractions ranging from 20% to 100% deuteration in 10% increments, omitting a 30% sample. The first excitonic peak as measured by UV-Vis spectroscopy is at 1.31 eV or a wavelength of 945 nm with a half-width at half-maximum of 58 nm. The nominal core sizes and polydispersity as measured by UV-Vis spectroscopy are 3.41 nm ($\sigma_d \leq 4.5\%$) according to the sizing curve reported by Maes *et al.*[3] with polydispersity calculated analogous to that of Weidman *et al.*[2] The full data set for these QDs is shown in Fig. 4. Solvent blanks were used to subtract background from each sample, which is mostly dominated by the solvent incoherent background. The q-independent background remaining comes from the QDs' core and ligand-shell incoherent background. As such, the remaining background should be the same across deuteration fractions since the same batch is used for each sample. The small dispersity allows 2 clear peaks in intensity to be captured in the neighborhood of 0.16 Å^{-1} and 0.26 Å^{-1}. As solvent deuteration fraction is increased, the first peak moves slightly toward larger q values.

Using less deuterated toluene decreases the SLD of the solvent and reduces contrast with the QDs, decreasing the measured intensity. There is a monotonic decrease of $I(q \to 0)$. When the solvent SLD matches that of the core, all scattering intensity will be due to contrast with the shell. Similarly, if the solvent SLD could match that of the ligand tails, then all intensity would arise from contrast with the core and ligand heads. The SLD of the oleic acid ligand tails is below that of hydrogen-bearing toluene, so this is not achievable in the experiment. We would expect that the minimum scattering intensity for $q \to 0$ would occur at the solvent match point. For a simple core-shell model with sharp interfaces between core, shell, and solvent using one SLD value of bulk oleic acid, this condition is given by

$$V_c(\varrho_{core} - \varrho_{solvent}) + V_s(\varrho_{shell} - \varrho_{solvent}) = 0, \tag{19}$$

where V_c is the volume of the core and V_s is the volume based on the distance from the QD center to the outside of the ligand shell. Equation (19) predicts the solvent match SLD to lie between the SLDs of the core and shell, with weighting toward one of the extremes based on the volume of each material. This would predict $I(q \to 0)$ to decrease as solvent SLD is decreased to the match point. If solvent SLD could be decreased more, then $I(q \to 0)$ would start increasing again. For typical QD sizes and materials used, the match condition, and thus minimum scattering intensity, will occur at a solvent SLD below that achievable with toluene. The match condition is not as straight forward for the composite model with solvent penetration. However, $I(q \to 0)$ remains monotonic with solvent SLD, indicating the solvent match condition is outside the range of deuteration fractions used.

Core-only and core-shell models are incapable of describing the data across all deuteration fractions. As demonstrated in Fig. S4 in the supplementary material, a core-only model incorrectly estimates the magnitude of scattering intensity at intermediate and low deuteration fractions. A core-shell model does not capture the low-q region at low deuteration fractions. These models are not sufficiently flexible enough to allow an unknown size and shape QD to be quantified from SANS data.

Figure 5 gives the MCMC fit to the SANS data using the MD-inspired model. In examining Fig. 5(b), a clear peak at the surface from ligand head groups is evident. Varying solvent deuteration fraction enables separation of the PbS and ligand contributions to the scattering. The high ligand volume fraction in the shell mirrors the high ligand coverage samples from MD simulation. Notably, despite this high ligand volume fraction, there is still a solvent presence near the QD surface, though a solvent peak is absent. This is consistent with facet-specific ligand binding leaving void space over the (100) core facets. However, in this size range, the core shapes have little exposed (100) surface area. This physical insight has implications for nanocrystal self-assembly where ligand coverage affects QD interactions and may direct orientated alignment of neighboring QDs.[18] There is also a sharp dropoff in the ligand shell when the extent of the ligand shell is reached and only solvent is present. Physically, this is sensible as ligands become extended at high ligand coverage, resulting in a largely spherical layer with a sharp transition to the solvent phase. In Figs. 5(a) and 5(b), multiple parameter sets sampled from the Markov chains are plotted as transparencies to give a sense of the spread in the parameter distributions. The dark saturation shows that many sampled models overlap and demonstrate a high confidence in the parameters. The median of each individual parameter distribution is used to plot the composite SLD profile in Fig. 5(b). Additionally, we separate the parameter set that best predicts the data according to the likelihood function. This parameter set is termed the most likely estimator (MLE). This composite SLD profile is virtually indistinguishable from the set of median parameter values. Figure 5(c) shows the fit of the SLD profile to the experimental data. The SLD profile predicts the intensity profiles well over the full q range and across all solvent deuteration fractions. The shift in the first q peak is captured because the fit SLD profile in the vicinity of the QD surface changes with deuteration fraction.

Although the MCMC method performance is invariant of the number of parameters included, caution must be exercised in choosing the appropriate model. The composite model used includes 8 parameters. In fitting any scattering data, model complexity must be considered to prevent overfitting. Statistical methods exist to consider model complexity in performing the fit. These methods penalize the fit according to the number of parameters. Two such methods, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), are used in this study.[56,57] The AIC is given by

$$\mathrm{AIC}(\boldsymbol{\theta}_{MLE}) = 2k - 2\ln(\mathcal{L}(\boldsymbol{\theta}_{MLE}|\mathbf{I})), \tag{20}$$

where $\boldsymbol{\theta}_{MLE}$ is the MLE parameter set, k is the number of parameters, and $\mathcal{L}$ is the normalized likelihood function. A better performing model will have a higher likelihood, but this is penalized by adding twice the number of parameters. The preferred model from a set is the one with the minimum AIC. The relative likelihood for each model $\exp\left(\frac{\mathrm{AIC}_{min}-\mathrm{AIC}_i}{2}\right)$ is proportional to the probability that the ith model best describes the data. These may be normalized across all the models by dividing by the sum of relative likelihoods to find the relative probability, or weight w_i, that model i is the best model of the set considered. The BIC penalizes the number of parameters more strongly,

$$\mathrm{BIC}(\boldsymbol{\theta}_{MLE}) = \ln(n)k - 2\ln(\mathcal{L}(\boldsymbol{\theta}_{MLE}|\mathbf{I})), \tag{21}$$

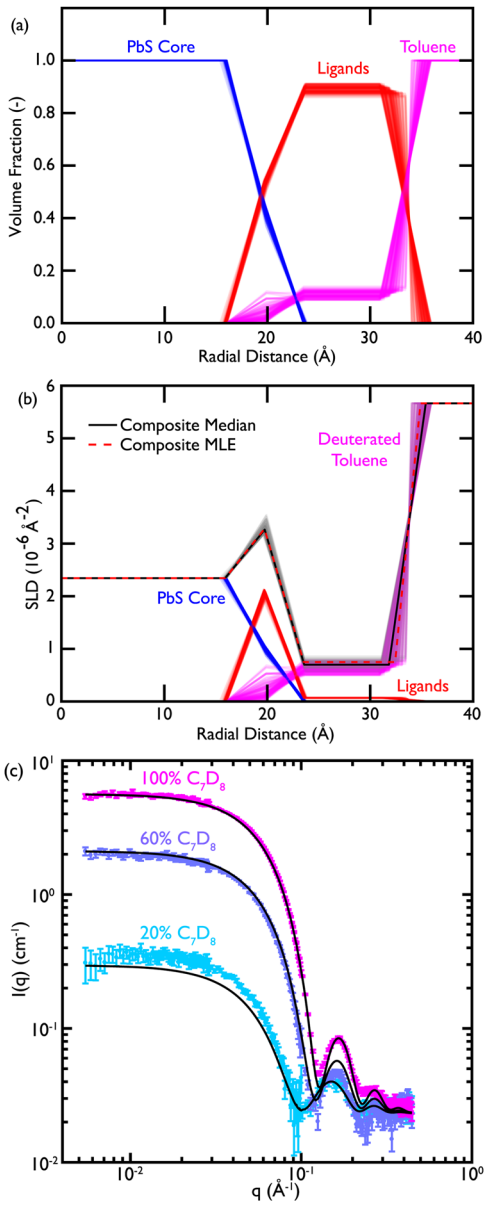

FIG. 5. MCMC fit to the SANS data of the MD-inspired model. 100 transparent walker samples are plotted for each series, demonstrating the credible region. (a) Fit volume fraction profile. (b) Fit SLD profile for a fully deuterated sample using the median (black) and MLE (red) of each parameter value. (c) Fit intensity profiles of 3 deuteration fractions chosen for clarity. The full fit is shown in the supplementary material.

TABLE I. Comparison of form factor models using the AIC and BIC to penalize model complexity based on the number of parameters.

Model	AIC_i	w_i	BIC_i	ΔBIC_i
Core-shell	51 172	0	51 199	54 915
Composite ($k = 7$)	−3 754	0.7	−3 716	0
Composite ($k = 8$)[a]	−3 753	0.3	−3 709	7

[a]Form factor described in the main text. See the supplementary material for a schematic of the other composite model.

with n being the number of data points. The difference between the minimum BIC and each individual BIC of a set may be used to assess the strength of evidence against models with higher BIC values. A difference of 2–6 is positive evidence, 6–10 is strong evidence, and more than 10 is said to be very strong evidence.[58] The AIC, AIC weights, BIC, and differences of BIC are given in Table I for the core-shell model and 2 separate composite models with varying degrees of complexity. The form factor presented in the main text uses 8 parameters. The other composite model is less complex with 7 parameters and does not allow a solvent peak at the surface. A schematic is given in the supplementary material. Both the AIC and BIC indicate that the 7 parameter model is the best descriptor of the data. The AIC, which does not penalize complexity as strongly, gives a probability of 0.7 that the 7 parameter model is optimal. The BIC shows that the 7 parameter model is slightly preferable, with a difference in BIC of 7. This would be considered strong evidence that the extra parameter is not necessary to describe the data. However, the fit results for this individual QD batch had a high ligand coverage such that most of the solvent was excluded at the surface and the solvent profile happened to be almost linear near the core surface. It is reasonable to expect that a sample with lower ligand coverage would have a larger solvent population at the surface as more solvent would penetrate the shell. As such, the 8 parameter model is selected to maintain flexibility across samples. The AIC concluded there is still a probability of 0.3 that the 8 parameter model is the best descriptor of the data. The 8 parameter model is a minimal model needed to resolve solvent penetration in the region near the QD surface separate from the ligand head groups. Any composite model which accounts for solvent penetration outperforms the core-shell model.

C. Analysis of parameter posteriors

1. Marginal and pairwise distributions

One of the advantages of the MCMC approach is the insight gained by examining the posterior distributions of the parameters. Figure 6 shows a corner plot of the parameters: the individual and pairwise distributions of the samples from the Markov chains. Polydispersity was not explicitly included in the model. In an ensemble measurement such as SANS, all of the individual QDs' intensity profiles are reflected in the integrated intensity profile. Populations with different sizes have proportional weight in the observed total intensity. A highly polydisperse sample is not suitable for fitting with this method without explicit inclusion of a polydispersity function. The omission of an explicit polydispersity function is permissible because the dominant source of smearing in this experiment comes from instrument resolution. If this were not the case, the acceptable parameter samples would not reflect the relative populations of those parameters in the experimental sample as the assumption that the data are normally distributed about the model would breakdown. We expect core polydispersities below ~5% to be suitable for omission of a polydispersity function. Smearing from instrument resolution dominates the total smearing of the model in this case. The supplementary material contains a discussion of this. This Bayesian method accounts for parameter polydispersity without assuming an underlying form (Gaussian, Schultz, etc.). For example, a bimodal distribution of core size may have resulted from the model fit if both populations describe the data equally well. Moreover, we obtain these distributions for all parameters, not only gross parameters such as core and shell size. Figure 6 demonstrates that symmetric distributions do describe some parameters such as f_{OA} or ϕ. However, the parameter r_{min} describing the distance to the closest facet of the QD PbS core is asymmetric and perhaps shows a second solution in the near vicinity. The marginal distribution is skewed to lower values. The median and MLE parameters are similar for this data set but are not identical. As these distributions describe how well the parameter can explain the data, certainty estimates are essential. The 95% credible region describes a range within which there is a 95% probability the parameter falls. This is the Bayesian equivalent of a confidence interval. Because an uninformative prior was used in this study, the resulting credible region gives the loosest bounds on the likely value of a parameter. This is akin to assuming nothing about the parameters ahead of time. If prior information is included from a separate measurement such as UV-Vis spectroscopy, then the resulting credible region may be tighter.

Aside from the marginal distributions of individual parameters, this method also yields correlations between pairs of parameters. Figure 6 shows correlation between the parameters $t_{plateau}$, t_{max}, f_{OA}, and ϕ. This can be interpreted as an exchange of one parameter for the other. A smaller $t_{plateau}$ is acceptable if t_{max} is made larger. The decay of the ligand profile acts as a pivot or lever about some marginal value. Accordingly, if the total amount of ligand changes, the total volume fraction scales accordingly to still fit the data. The pairwise distributions are also useful in the model iteration procedure. Strong correlations between parameters direct the researcher to probe the model for ill-posed behavior and eliminate redundant parameters.

2. Complete characterization

The combination of MD simulation and fitting via MCMC provides powerful resolution into each material separately. The shape factor is posed as a sum of each material with the relatively larger number of parameters needed to do so estimated with an efficient algorithm. Each individual volume fraction profile $\varphi(r)$ may then be analyzed. The total mass of each material is given by the integrated mass density. This is obtained by radial integration of the volume fraction profile and scaling by the bulk density,

$$m_i = \int_V \rho(r)dV = \int_0^\infty 4\pi r^2 \rho_i^{\text{bulk}} \varphi_i(r)dr, \tag{22}$$

with m_i being the mass of each material, $\rho(r)$ being the mass density profile, ρ_i^{bulk} being the bulk density of material i, and $\varphi_i(r)$ being the fit volume fraction profile of each material. The number of molecules

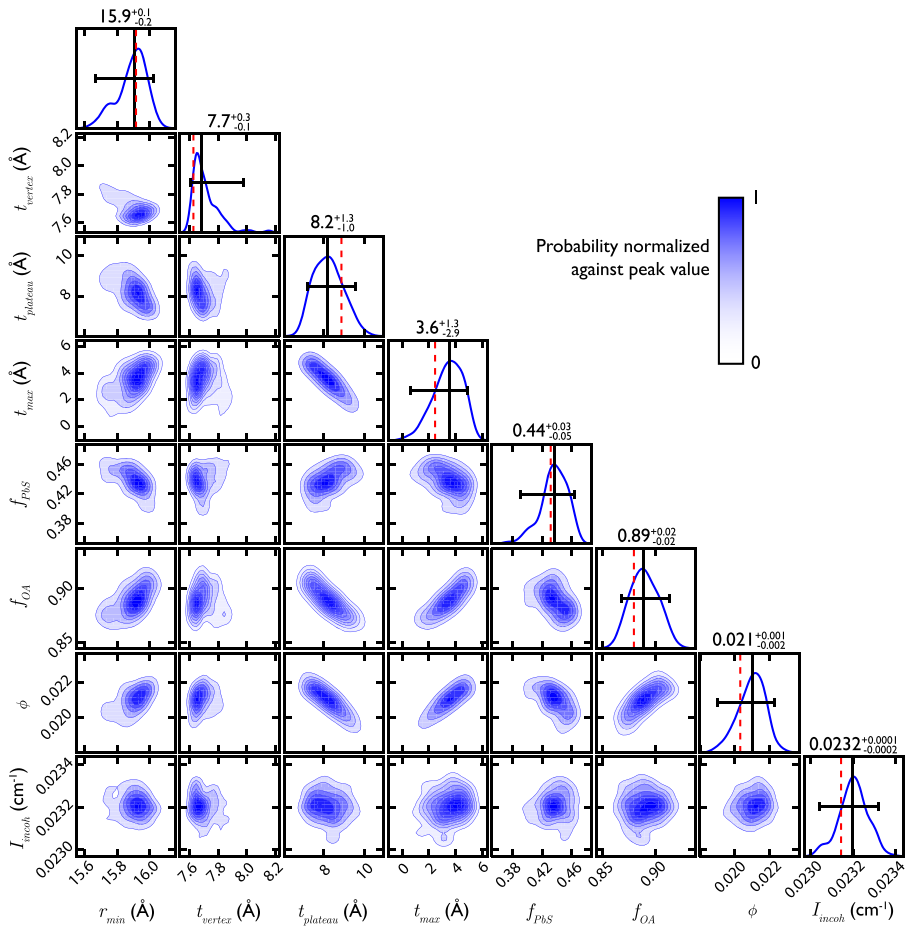

FIG. 6. Corner plot of fit parameters. The main diagonal shows the marginal distribution of each parameter integrating out all others (blue), distribution median (black, vertical), center 95% credible region (black, horizontal), and most likely estimator (red, dotted). Values above each column specify the median and credible region. Off diagonal plots show the correlation between each pair of parameters. Circular distributions show no correlation, while diagonal distributions show strong correlations. Each pairwise distribution is normalized by the peak value.

of each material is then found using the molecular weight and Avogadro's number: $N_i = \frac{m_i N_A}{MW_i}$. This methodology is also applied during fitting to ensure the number of ligand head groups equals the number of full oleate molecules bound to the nanocrystal.

In the bulk, PbS has a stoichiometric ratio of 1:1. As nanocrystals, we expect the Pb:S ratio to be in the range of 1.1–1.3 for a truncated octahedron core. The number of paired lead and sulfur atoms calculated from the mass density integration must equal half the sum of the number of Pb and S atoms in the core. Additionally, the core and ligands must be charge balanced. We assume that each oleate ligand supplies a 1− charge and each excess lead atom supplies a 2+ charge. Under these assumptions, the Pb:S ratio is shifted to a value that satisfies mass and charge balance according to the fit $\varphi(r)$ profiles. We solve for $N_{Pb,core}$ and $N_{S,core}$ using

$$N_{Pb,core} + N_{S,core} = 2N_{PbS}, \tag{23}$$

$$2(N_{Pb,core} - N_{S,core}) - N_{ligand} = 0. \tag{24}$$

In addition to the atomic ratio, we calculate the equivalent spherical volume diameter of the core. The equivalent spherical diameter is found from a sphere with the density of PbS containing the same mass of PbS as calculated from the posteriors. Ligand coverage is computed by dividing the total number of ligand molecules by the surface area. The surface area used for this calculation is based on the equivalent spherical diameter and divided by 0.9212, the maximal sphericity of a truncated octahedron family QD. Table II contains a summary of this complete characterization including the nominal size as measured by UV-Vis spectroscopy.[3] The values given represent the median of these computed descriptors for 100 samples from the posteriors as well as the central 95% confidence region. The measured core size is larger than the UV-Vis sizing curve result. Maes *et al.* cite up to 10% errors in their predicted core sizes at this size as opposed to smaller error for larger cores.[3] There is also systematic error resulting from the particular discretization of the core surface with one point at $\frac{1}{2}t_{vertex}$. The model cannot resolve a smoother shape of the ligand volume fraction profile a few Angstroms outside of r_{min}. We could have fixed the ligand head peak near the surface at $\frac{2}{3}t_{vertex}$, and this would reduce the core size to keep the peak location the same. Attempts at using more parameters to resolve the peak location and ligand profile did yield smaller core sizes of ~3.6 nm, but this is still larger than the 3.41 nm result from UV-Vis. Given these considerations, we selected the minimal model described previously. Increased model resolution is not justifiable within the framework described.

TABLE II. Summary of QD characterization. Values given represent the median of the computed descriptors from 100 samples of the posterior. Bounds denote the extents of the central 95% credible region.

UV-Vis nominal size (nm)	3.41
Equivalent volume diameter (nm)	$3.95^{3.97}_{3.93}$
Number of ligands per QD	199^{203}_{195}
Ligand coverage (nm^{-2})	$4.06^{4.10}_{4.03}$
Pb (%)	$54.4^{54.4}_{54.3}$
S (%)	$45.6^{45.7}_{45.6}$
Core Pb:S	$1.192^{1.194}_{1.190}$

The measured Pb/S ratio is within the expected range, albeit it is slightly lower than a value of ~1.3 reported previously for QDs of this size.[55] This is likely due to the manner of calculating the atomic ratio from the ligand charge balance constraint. We have measured the number of ligands per QD using quantitative NMR as 222 ± 3 based on the vinyl protons' signal with a dibromomethane standard.[27] The error in this value is quoted from measurement repetitions and does not account for error in the measured number density of QDs from UV-Vis spectroscopy. As such, the uncertainty is likely larger. This number is only slightly higher than the number of ligands per QD measured with SANS, with an error of 10.4%. The measured ligand coverage agrees well with previous experimental measurements on PbS and PbSe in the range of 3–5 nm^{-2}.[18,50,59] This is a self-consistent result that also matches literature results. The largest source of error stems from the necessity of parsing the ligand head and tail groups separately. With a single technique, we obtain a complete structural characterization of the QDs with rigorous treatment of the experimental uncertainty and resolution afforded by the technique.

IV. EXTENSION TO OTHER NANOCRYSTALS

The same methodology presented in this study may be extended to other colloidal nanocrystal systems. In designing a set of experiments to completely characterize a batch of nanocrystals, the SLDs of all relevant materials including the core, shell, ligand, and solvent must be compared to ensure sufficient resolution. If the core and ligands have similar scattering length densities, then it will be difficult to resolve them separately. Table III contains the SLDs of various nanocrystal materials, including cubic perovskite QDs,[60–63] and some typical solvents. Toluene spans the SLD range from 0.94 to 5.66 × 10^{-6} $Å^{-2}$, and mixed hexanes span the range from −0.57 to 6.14 × 10^{-6} $Å^{-2}$ as deuteration fraction is increased. Solvent contrast with the nanocrystals can be maximized by using the largest difference between nanocrystal SLD and solvent SLD. Binary components may also be considered. Of the lead and cadmium chalcogenides, all core-shell mixtures can be resolved separately except for PbS and CdSe which have nearly identical SLDs. Cesium-based perovskites have comparable SLDs, though methylammonium(MA)-based perovskites have lower SLDs. In addition to solvent contrast, the expected incoherent background scattering should be estimated to ensure enough signal will be collected. Neglecting multiple scattering events, a good estimate for the incoherent background term is

TABLE III. Relevant material specific quantities for SANS measurements of various nanocrystals.

Material	SLD (10^{-6} $Å^{-2}$)	Σ_c (cm^{-1})	Σ_i (cm^{-1})	Σ_c/Σ_i
PbS	2.34	0.18	1.9×10^{-4}	943
PbSe	2.96	0.32	0.055	58.5
CdS	1.54	0.074	0.070	1.06
CdSe	2.34	0.19	0.069	2.73
InP	1.83	0.11	0.011	9.80
Au	4.66	0.46	0.025	18.3
Ag	3.47	0.26	0.034	7.60
$CsPbBr_3$	1.78	0.16	0.003	60.7
$CsPbI_3$	2.72	0.095	0.005	20.7
$MAPbBr_3$	0.22	0.005	0.46	0.012
$MAPbI_3$	0.46	0.009	1.20	0.008
Toluene	0.94	0.013	3.6	3.6×10^{-3}
d_8-toluene	5.66	0.47	0.093	5.1
Hexanes	−0.57	4.5×10^{-3}	5.1	8.7×10^{-4}
d_8-hexanes	6.14	0.51	0.13	3.9
Oleic acid	0.078	7.4×10^{-5}	5.2	1.4×10^{-5}

the macroscopic incoherent scattering cross section Σ_i, given by

$$\Sigma_i = \frac{4\pi b_i^2 \rho N_A}{MW}, \tag{25}$$

where b_i is the incoherent scattering length, ρ is the density, N_A is Avogadro's number, and MW is the molecular weight. When multiple materials are used, the total background term is the volume fraction weighted sum of incoherent macroscopic scattering cross sections of all the materials. This background term should be compared to the macroscopic coherent scattering cross section Σ_c to gauge the signal-to-noise ratio. Σ_c is defined analogously to Eq. (25) with the coherent scattering length b_c in place of b_i. These values for various nanocrystals and solvents are also given in Table III. It should be noted that the total signal is not given by the coherent macroscopic scattering cross section. A representative form factor and volume fraction of scattering particles would need to be chosen and calculated to estimate the observed signal. The example system of PbS used in this study has the highest signal-to-noise ratio (Σ_c/Σ_i). The lowest signal-to-noise ratio of traditional nanocrystal materials is that of CdS. However, this methodology should still work for CdS. For a ~1.0 v/v% solution in toluene, i.e., dilute, the incoherent background term after subtracting off the solvent blank should be ~0.03 cm^{-1} with most of this background coming from the ligands. Comparing this to the experimental data of this study in Fig. 4, the first peak will still be above the background. Moreover, CdS would experience more contrast with the solvent due to a larger difference between SLDs. Similarly for the perovskites, cesium-based perovskite QDs will be amenable to SANS characterization. Methylammonium-based perovskites are likely infeasible for SANS characterization due to the low signal-to-noise ratio. Following this analysis, all common nanocrystal solutions can be characterized using this methodology except methylammonium-based perovskite QDs.

V. CONCLUSION

Here, we have demonstrated the power of SANS in fully characterizing nanocrystal solutions. This study outlines the methodology of using a single technique to extract size, shape, ligand coverage, and concentration from SANS. The approach may be extended to most nanocrystal systems. We project the utility of this work for studying both core and ligand-shell structure in single or multicomponent systems. Resolution to each material arose from posing the shape factor as a composite of separate volume fraction profiles. This model was inspired by examining the structure of single-QD MD simulations. This is possible because the system is well-understood and so can be simulated with high fidelity. Solvent variation enhanced resolution in the fitting. By varying solvent deuteration fraction, the environment near the QD surface may be parsed separately from the local solvent population. It is necessary to account for ligand head groups localized at the nanocrystal surface. Although more parameters are needed in a model like this, the MCMC approach efficiently fits the data. The algorithm itself takes advantage of parallel implementation but is accelerated by having an analytical expression for the intensity profile, implementing instrument smearing with accurate numerical quadrature, and omitting polydispersity convolution integrals. This study serves as validation for the affine invariant MCMC method as an effective data fitting tool for any Bayesian parameter estimation. In the neutron scattering community, the method may be applied to other systems to both accelerate form factor selection and elucidate minute structural details, which have profound impact on material performance with unprecedented fidelity.

SUPPLEMENTARY MATERIAL

See supplementary material for the following files. supp_info.pdf—Contains NMR spectra used for ligand quantitation, additional core shape MD volume fraction profiles, comparison to core-only and core-shell models with and without polydispersity, the full fit to the data, and an alternate composite form factor model schematic. sans_data.csv—Spreadsheet of background corrected SANS data including q resolution and scattering intensity at eight deuteration fractions. fit_data_mcmc.m—MATLAB script fitting the SANS data with the MCMC approach. External functions and data structure used for fitting are included with the script.

ACKNOWLEDGMENTS

The authors thank B. Olsen for helpful conversations in interpreting SANS data. Access to NGB30mSANS was provided by the Center for High Resolution Neutron Scattering, a partnership between the National Institute of Standards and Technology and the National Science Foundation under Agreement No. DMR-1508249. Certain commercial equipment, instruments, or materials are identified in this document. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products identified are necessarily the best available for the purpose. W.A.T. was supported by the Camille and Henry Dreyfus Foundation. J.W.S. was supported by the MIT Portugal Seed Fund.

APPENDIX A: DERIVATION OF INTENSITY PROFILE FROM A PIECEWISE LINEAR SLD PROFILE

This study assumes a piecewise linear SLD profile, given by Eq. (5) in Sec. II D 1. For computing the form factor from this profile, it is useful to define the following variables, spherical Bessel functions, and integral identities:

$$\alpha_1 = \frac{r_1}{r_2 - r_1}, \tag{A1}$$

$$\alpha_2 = \frac{r_2}{r_2 - r_1}, \tag{A2}$$

$$\alpha_2 - \alpha_1 = 1, \tag{A3}$$

$$j_0(x) = \frac{\sin(x)}{x}, \tag{A4}$$

$$j_1(x) = \frac{\sin(x) - x\cos(x)}{x^2}, \tag{A5}$$

$$y_0(x) = -\frac{\cos(x)}{x}, \tag{A6}$$

$$y_1(x) = -\frac{\cos(x) + x\sin(x)}{x^2}, \tag{A7}$$

$$\int r^2 j_0(qr)dr = r^3 \frac{j_1(qr)}{qr} + C, \tag{A8}$$

$$\int r^3 j_0(qr)dr = r^4\left(\frac{j_1(qr)}{qr} - \frac{y_1(qr)}{(qr)^2} - \frac{y_0(qr)}{(qr)^3}\right) + C. \tag{A9}$$

The derivation begins from the definition of the form factor in Eq. (3) with a linear SLD profile from Eq. (5) between two points (r_1, ρ_1) and (r_2, ρ_2),

$$\frac{F(q)}{4\pi} = \int_{r_1}^{r_2} (\rho_1 - \rho_{solv}) r^2 j_0(qr)dr + \int_{r_1}^{r_2}\left(\frac{\rho_2 - \rho_1}{r_2 - r_1}\right) r^3 j_0(qr)dr - \int_{r_1}^{r_2}\left(\frac{\rho_2 - \rho_1}{r_2 - r_1}\right) r_1 r^2 j_0(qr)dr. \tag{A10}$$

Each term is evaluated with the integral identities defined above and expanded,

$$\begin{aligned}\frac{F(q)}{4\pi} = {} & (\rho_1 - \rho_{solv})\left(r_2^3\frac{j_1(qr_2)}{qr_2} - r_1^3\frac{j_1(qr_1)}{qr_1}\right) \\ & + \left(\frac{\rho_2 - \rho_1}{r_2 - r_1}\right) r_2^4\left(-\frac{y_1(qr_2)}{(qr_2)^2} - \frac{y_0(qr_2)}{(qr_2)^3}\right) \\ & - \left(\frac{\rho_2 - \rho_1}{r_2 - r_1}\right) r_1^4\left(-\frac{y_1(qr_1)}{(qr_1)^2} - \frac{y_0(qr_1)}{(qr_1)^3}\right) \\ & + \left(\frac{\rho_2 - \rho_1}{r_2 - r_1}\right)\left(r_2^4\frac{j_1(qr_2)}{qr_2} - r_1^4\frac{j_1(qr_1)}{qr_1}\right) \\ & - \left(\frac{\rho_2 - \rho_1}{r_2 - r_1}\right) r_1 r_2^3\frac{j_1(qr_2)}{qr_2} + \left(\frac{\rho_2 - \rho_1}{r_2 - r_1}\right) r_1 r_1^3\frac{j_1(qr_1)}{qr_1}.\end{aligned} \tag{A11}$$

We apply the definitions of α_1 and α_2, simplify terms, and convert radii to volumes to arrive at

$$F(q) = 3(\rho_2 - \rho_{solv})V_2\left(\frac{j_1(qr_2)}{qr_2}\right) - 3(\rho_1 - \rho_{solv})V_1\left(\frac{j_1(qr_1)}{qr_1}\right) + 3(\rho_2 - \rho_1)\alpha_2 V_2\left(-\frac{y_1(qr_2)}{(qr_2)^2} - \frac{y_0(qr_2)}{(qr_2)^3}\right) - 3(\rho_2 - \rho_1)\alpha_1 V_1\left(-\frac{y_1(qr_1)}{(qr_1)^2} - \frac{y_0(qr_1)}{(qr_1)^3}\right). \quad \text{(A12)}$$

This expression is not well-posed for the point $r = 0$ due to the spherical Bessel functions of the first kind. In the model, we assume that the PbS profile is flat in the first interval with $r_1 = 0$ to $r_2 = r_{min}$. The expression for a spherical scatterer is applied in this region. The expression is also undefined when $r_1 = r_2$. In this case, the expression should reduce to the form factor of a core-shell model. By taking the limit as r_2 approaches r_1, the core-shell model is recovered.

APPENDIX B: INSTRUMENT SMEARING USING GAUSS-HERMITE QUADRATURE

Gauss-Hermite quadrature is suitable for approximating the value of integrals of the form

$$g(x) = \int_{-\infty}^{\infty} \exp(-x^2) f(x) dx \approx \sum_{j=1}^{n} w_j f(x_j). \quad \text{(B1)}$$

Here, x_j are the roots of the Hermite polynomial $H_n(x_j)$ with the associated weights,

$$w_j = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(x_j)]^2}. \quad \text{(B2)}$$

The resolution function for each measured q_i is Gaussian distributed and known from calibration,

$$R(q; \bar{q}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(q - \bar{q})^2}{2\sigma^2}\right). \quad \text{(B3)}$$

The smeared intensity is calculated using

$$I_{smear}(q_i) = \int_{-\infty}^{\infty} R(q'; \bar{q}_i, \sigma_i) I(q') dq'. \quad \text{(B4)}$$

The argument of the exponent in the resolution function is used to define the x from Eq. (B1),

$$x = \frac{q' - \bar{q}_i}{\sqrt{2}\sigma_i}. \quad \text{(B5)}$$

Using this, the change of variables is made to yield the integral

$$I_{smear}(q_i) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-x^2) I\left(\sqrt{2}\sigma_i x + \bar{q}_i\right) dx. \quad \text{(B6)}$$

Gauss-Hermite quadrature is applied such that we implement

$$I_{smear}(q_i) \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{n} w_j I\left(\sqrt{2}\sigma_i x_j + \bar{q}_i\right). \quad \text{(B7)}$$

It is found that 12 terms in the sum is sufficient for convergence. In practice, the algebraic expression for $I(q)$ is evaluated not at the measured q_i but at $\left(\sqrt{2}\sigma_i x_j + \bar{q}_i\right)$ for each x_j, weighted by w_j and summed up to find the observed, smeared value of $I(q_i)$ for each measured q_i.

REFERENCES

[1] I. Moreels, K. Lambert, D. Smeets, D. De Muynck, T. Nollet, J. C. Martins, F. Vanhaecke, A. Vantomme, C. Delerue, G. Allan, and Z. Hens, ACS Nano **3**, 3023 (2009).
[2] M. C. Weidman, M. E. Beck, R. S. Hoffman, F. Prins, and W. A. Tisdale, ACS Nano **8**, 6363 (2014).
[3] J. Maes, N. Castro, K. De Nolf, W. Walravens, B. Abécassis, and Z. Hens, Chem. Mater. **30**, 3952 (2018).
[4] Y. Yang, H. Qin, M. Jiang, L. Lin, T. Fu, X. Dai, Z. Zhang, Y. Niu, H. Cao, Y. Jin, F. Zhao, and X. Peng, Nano Lett. **16**, 2133 (2016).
[5] M. C. Weidman, K. G. Yager, and W. A. Tisdale, Chem. Mater. **27**, 474 (2014).
[6] A. P. Alivisatos, Science **271**, 933 (1996); e-print arXiv:1011.1669v3.
[7] A. J. Nozik, Physica E **14**, 115 (2002).
[8] G. Konstantatos, I. Howard, A. Fischer, S. Hoogland, J. Clifford, E. Klem, L. Levina, and E. H. Sargent, Nature **442**, 180 (2006).
[9] G. I. Koleilat, L. Levina, H. Shukla, S. H. Myrskog, S. Hinds, A. G. Pattantyus-Abraham, and E. H. Sargent, ACS Nano **2**, 833 (2008).
[10] J. J. Choi, Y. F. Lim, M. B. Santiago-Berrios, M. Oh, B.-R. Hyun, L. Sun, A. C. Bartnik, A. Goedhart, G. G. Malliaras, H. D. Abruña, F. W. Wise, and T. Hanrath, Nano Lett. **9**, 3749 (2009).
[11] D. V. Talapin, J.-S. Lee, M. V. Kovalenko, and E. V. Shevchenko, Chem. Rev. **110**, 389 (2010).
[12] M. T. Frederick, V. A. Amin, N. K. Swenson, A. Y. Ho, and E. A. Weiss, Nano Lett. **13**, 287 (2013).
[13] M. S. Azzaro, A. Dodin, D. Y. Zhang, A. P. Willard, and S. T. Roberts, Nano Lett. **18**, 3259 (2018).
[14] D. Bozyigit, N. Yazdani, M. Yarema, O. Yarema, W. M. M. Lin, S. Volk, K. Vuttivorakulchai, M. Luisier, F. Juranyi, and V. Wood, Nature **531**, 618 (2016).
[15] N. Yazdani, D. Bozyigit, K. Vuttivorakulchai, M. Luisier, I. Infante, and V. Wood, Nano Lett. **18**, 2233 (2018).
[16] W. D. Luedtke and U. Landman, J. Phys. Chem. **100**, 13323 (1996).
[17] K. Bian, J. J. Choi, A. Kaushik, P. Clancy, D. M. Smilgies, and T. Hanrath, ACS Nano **5**, 2815 (2011).
[18] J. J. Choi, C. R. Bealing, K. Bian, K. J. Hughes, W. Zhang, D.-M. Smilgies, R. G. Hennig, J. R. Engstrom, and T. Hanrath, J. Am. Chem. Soc. **133**, 3131 (2011).
[19] A. P. Kaushik and P. Clancy, J. Chem. Phys. **136**, 114702 (2012).
[20] Z. Wang, C. Schliehe, K. Bian, D. Dale, W. A. Bassett, T. Hanrath, C. Klinke, and H. Weller, Nano Lett. **13**, 1303 (2013).
[21] M. P. Boneschanscher, W. H. Evers, J. J. Geuchies, T. Altantzis, B. Goris, F. T. Rabouw, S. A. P. van Rossum, H. S. J. van der Zant, L. D. A. Siebbeles, G. V. Tendeloo, I. Swart, J. Hilhorst, A. V. Petukhov, S. Bals, and D. Vanmaekelbergh, Science **344**, 1377 (2014).
[22] B. W. Goodfellow, Y. Yu, C. A. Bosoy, D. M. Smilgies, and B. A. Korgel, J. Phys. Chem. Lett. **6**, 2406 (2015).
[23] M. C. Weidman, D.-M. Smilgies, and W. A. Tisdale, Nat. Mater. **15**, 775 (2016).
[24] M. C. Weidman, Q. Nguyen, D.-M. Smilgies, and W. A. Tisdale, Chem. Mater. **30**, 807 (2018).
[25] J. S. Owen, J. Park, P.-E. Trudeau, and A. P. Alivisatos, J. Am. Chem. Soc. **130**, 12279 (2008).
[26] B. Fritzinger, R. K. Capek, K. Lambert, J. C. Martins, and Z. Hens, J. Am. Chem. Soc. **132**, 10195 (2010).
[27] N. C. Anderson, M. P. Hendricks, J. J. Choi, and J. S. Owen, J. Am. Chem. Soc. **135**, 18536 (2013).
[28] K. Levenberg, Q. Appl. Math. **2**, 164 (1944); e-print arXiv:1011.1669v3.
[29] D. W. Marquardt, J. Soc. Ind. Appl. Math. **11**, 431 (1963).
[30] N. Christensen, R. Meyer, L. Knox, and B. Luey, Classical Quantum Gravity **18**, 2677 (2001); e-print arXiv:0103134 [astro-ph].
[31] J. Dunkley, M. Bucher, P. G. Ferreira, K. Moodley, and C. Skordis, Mon. Not. R. Astron. Soc. **356**, 925 (2005); e-print arXiv:0405462 [astro-ph].
[32] J. Goodman and J. Weare, Commun. Appl. Math. Comput. Sci. **5**, 65 (2010).
[33] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125**, 306 (2013).

[34]J. Akeret, S. Seehars, A. Amara, A. Refregier, and A. Csillaghy, Astron. Comput. **2**, 27 (2013); e-print arXiv:1212.1721.
[35]J. Vrugt and C. Ter Braak, Hydrol. Earth Syst. Sci. **15**, 3701 (2011).
[36]I. Bressler, B. R. Pauw, and A. F. Thünemann, J. Appl. Crystallogr. **48**, 962 (2015).
[37]E. Drijvers, J. De Roo, J. C. Martins, I. Infante, and Z. Hens, Chem. Mater. **30**, 1178 (2018).
[38]W. Paul, D. Y. Yoon, and G. D. Smith, J. Chem. Phys. **103**, 1702 (1995).
[39]C. D. Wick, M. G. Martin, and J. I. Siepmann, J. Phys. Chem. B **104**, 8008 (2000).
[40]P. Schapotschnikow, M. A. van Huis, H. W. Zandbergen, D. Vanmaekelbergh, and T. J. H. Vlugt, Nano Lett. **10**, 3966 (2010).
[41]M. Grünwald, A. Zayak, J. B. Neaton, P. L. Geissler, and E. Rabani, J. Chem. Phys. **136**, 234111 (2012); e-print arXiv:1201.2992.
[42]A. P. Kaushik and P. Clancy, J. Comput. Chem. **34**, 523 (2013).
[43]Z. Fan, R. S. Koster, S. Wang, C. Fang, A. O. Yalcin, F. D. Tichelaar, H. W. Zandbergen, M. A. Van Huis, and T. J. H. Vlugt, J. Chem. Phys. **141**, 244503 (2014).
[44]A. Widmer-Cooper and P. Geissler, Nano Lett. **14**, 57 (2014).
[45]J. A. Anderson, C. D. Lorenz, and A. Travesset, J. Comput. Phys. **227**, 5342 (2008).
[46]J. Glaser, T. D. Nguyen, J. A. Anderson, P. Lui, F. Spiga, J. A. Millan, D. C. Morse, and S. C. Glotzer, Comput. Phys. Commun. **192**, 97 (2015); e-print arXiv:1412.3387.
[47]T. D. Nguyen, C. L. Phillips, J. A. Anderson, and S. C. Glotzer, Comput. Phys. Commun. **182**, 2307 (2011).
[48]M. A. Hines and G. D. Scholes, Adv. Mater. **15**, 1844 (2003).
[49]A. E. Colbert, W. Wu, E. M. Janke, F. Ma, and D. S. Ginger, J. Phys. Chem. C **119**, 24733 (2015).
[50]I. Moreels, B. Fritzinger, J. C. Martins, and Z. Hens, J. Am. Chem. Soc. **130**, 15081 (2008).
[51]C. J. Glinka, J. G. Barker, B. Hammouda, S. Krueger, J. J. Moyer, and W. J. Orts, J. Appl. Crystallogr. **31**, 430 (1998).
[52]S. R. Kline, J. Appl. Crystallogr. **39**, 895 (2006).
[53]G. V. Damme, Legendre laguerre and hermite—Gauss quadrature, 2010, https://www.mathworks.com/matlabcentral/fileexchange/26737-legendre-laguerre-and-hermite-gauss-quadrature.
[54]A. Grinsted, Gwmcmc, 2015, https://github.com/grinsted/gwmcmc.
[55]H. Choi, J.-H. Ko, Y.-H. Kim, and S. Jeong, J. Am. Chem. Soc. **135**, 5278 (2013).
[56]H. Akaike, IEEE Trans. Autom. Control **19**, 716 (1974).
[57]G. Schwarz, Ann. Stat. **6**, 461 (1978); e-print arXiv:1011.1669v3.
[58]R. E. Kass and A. E. Raferty, J. Am. Stat. Assoc. **90**, 773 (1995).
[59]W.-L. Ong, S. M. Rupich, D. V. Talapin, A. J. McGaughey, and J. A. Malen, Nat. Mater. **12**, 410 (2013).
[60]J. Chen, D. J. Morrow, Y. Fu, W. Zheng, Y. Zhao, L. Dang, M. J. Stolt, D. D. Kohler, X. Wang, K. J. Czech, M. P. Hautzinger, S. Shen, L. Guo, A. Pan, J. C. Wright, and S. Jin, J. Am. Chem. Soc. **139**, 13525 (2017).
[61]W. Zhou, F. Sui, G. Zhong, G. Cheng, M. Pan, C. Yang, and S. Ruan, J. Phys. Chem. Lett. **9**, 4915 (2018).
[62]Y. Wang, X. Lü, W. Yang, T. Wen, L. Yang, X. Ren, L. Wang, Z. Lin, and Y. Zhao, J. Am. Chem. Soc. **137**, 11144 (2015).
[63]K. P. Ong, T. W. Goh, Q. Xu, and A. Huan, J. Phys. Chem. A **119**, 11033 (2015).