Postprint

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-33847

Massively parallel approximate Bayesian computation for estimating nanoparticle diffusion coefficients, sizes and concentrations using confocal laser scanning microscopy

Magnus Röding[1,*] and Markus Billeter[2]

[1] RISE Research Institutes of Sweden, Bioscience and Materials, Göteborg, Sweden

[2] Department of Space, Earth and Environment, Chalmers University of Technology, Göteborg, Sweden

* Corresponding author: `magnus.roding@ri.se`

Abstract

We implement a massively parallel population Monte Carlo approximate Bayesian computation (PMC-ABC) method for estimating diffusion coefficients, sizes and concentrations of diffusing nanoparticles in liquid suspension using confocal laser scanning microscopy (CLSM) and particle tracking. The method is based on the joint probability distribution of diffusion coefficients and the time spent by a particle inside a detection region where particles are tracked. We present freely available CPU and GPU versions of the analysis software, and we apply the method to characterize mono- and bidisperse samples of fluorescent polystyrene beads.

Keywords

1 Introduction

Functional nanoparticles, both natural and synthetic, are widely used within biological matter e.g. as tracer particles for biomedical imaging, as biomarkers for diagnostics (Chironi et al., 2009; Doeuvre et al., 2009; Khare and Dokholyan, 2007; Nune et al., 2009; Weissleder et al., 2014), and as biocarriers for targeted delivery of therapeutic agents (Filipe et al., 2011, 2012; Remaut et al., 2007; Yoo et al., 2011). In this context, not only surface and interaction properties of the particles but also size (distribution), concentration, and even shape matters for uptake, efficiency, and toxicity (Banerjee et al., 2016; Kumar et al., 2015; Salatin et al., 2015; Shang et al., 2014; Wilhelm et al., 2016). Therefore, there is a demand for detection and characterization methods for sub-micron particles,

including particles suspended in complex fluids like undiluted blood (Braeckmans et al., 2010; Montes-Burgos et al., 2010; Nel et al., 2009). There is a plethora of methods available for obtaining particle size and concentration in different contexts, including transmission electron microscopy, flow cytometry, and dynamic light scattering. Of particular relevance for this work, however, are particle tracking-based microscopy techniques.

A number of particle tracking-based microscopy techniques have been presented for characterization of nanoparticles suspended in solution (Braeckmans et al., 2010; Carr and Wright, 2008; Du et al., 2010; Röding et al., 2011, 2013a,b, 2016; Saveyn et al., 2010). By imaging their Brownian motion in suspension, particle parameters including size and concentration can be estimated. In previous work (Röding et al., 2011), we demonstrated for the first time that estimation of number concentration can be performed in a 'self-calibrated' fashion by jointly estimating the diffusion coefficient, the size of the detection region in which the particles can be detected and tracked, and the absolute number concentration. The size of the detection region is a function of field of view, depth of field, sample illumination, particle fluorescence intensity, and the particle detection and tracking algorithm. Its size and in particular its axial dimension (along the optical axis) is therefore not constant between samples and experimental setups; the novelty of that paper was the direct estimation of this size, and the method has been proven useful in a number of applications (Deschout et al., 2014; Forier et al., 2014; Naeye et al., 2013; Zagato et al., 2014). In later work (Röding et al., 2016), because closed-form modelling of this problem is difficult but simulation of the data-generating process is straightforward, we conclude that the problem naturally lends itself to simulation-based inference. Consequently, we introduced an approximate Bayesian computation (ABC)-based method for monodisperse particles.

In this work, we implement a massively parallel ABC method based on population Monte Carlo ABC (PMC-ABC) with implementations for both CPU and GPU. Beside being

significantly faster than the previous ABC method, we generalize its use to bidisperse particles. The method is based on the joint probability distribution of diffusion coefficients and the time spent by a particle inside the detection region. We demonstrate the usefulness of the method on mono- and bidisperse fluorescent polystyrene beads, using a commercially available confocal laser scanning microscope (CLSM).

2 Results and discussion

2.1 Problem setup

Consider a liquid suspension filling the volume $\Omega = [-A_x/2, A_x/2] \times [-A_y/2, A_y/2] \times [-A_z/2, A_z/2]$. In a real experiment, the lateral sizes A_x and A_y (perpendicular to the optical axis z) are in the order of millimeters, and the axial dimension A_z is in the order of hundreds of micrometers. The cuboidal shape is only a convenient assumption for implementing periodic boundary conditions in simulations and does not reflect the true shape. The liquid suspension is populated with freely diffusing, Brownian particles with diffusion coefficient D in diffusion equilibrium and a concentration (density) c, i.e. a Poisson distributed number $\text{Poi}(c\,|\Omega|)$ of particles are uniformly distributed in Ω at any time. A microscope is used to image and track particles in two dimensions inside a cuboid mid-section of the liquid suspension which we denote the detection region, $\omega = [-a_x/2, a_x/2] \times [-a_y/2, a_y/2] \times [-a_z/2, a_z/2]$. The lateral sizes a_x and a_y of the detection region are determined by the field of view of the microscope, and known before experiments are commenced. The axial size a_z (parallel to the optical axis z) depends on several experimental factors such as depth of field, laser intensity, pinhole size, particle fluorescence intensity, and the particle detection algorithm, and is unknown prior to data analysis. For a polydisperse particle system, the parameters D, a_z, and c are generalized to $D_1, D_2, ..., a_{z,1}, a_{z,2}, ...,$ and $c_1, c_2, ...$

4

A number of videos are recorded with time lag Δt between consecutive frames and particles are detected and tracked in two dimensions during the time they reside in the detection region ω. The number of positions k in the recorded trajectories vary between 1 and $k_{\max}$ (the number of frames in the longest video). In practice, we only keep trajectories with $k \geq k_{\min} = 3$ because short trajectories are often the result of false positives in the detection phase due to imaging noise (Jaqaman et al., 2008). Assuming a single diffusion coefficient D for now, each diffusing particle is subject to independent Gaussian displacements $\Delta x, \Delta y \sim \mathcal{N}(\mu = 0, \sigma^2 = 2D\Delta t)$ (and a third, unobserved displacement Δz with the same distribution) (Berg, 1993). The data obtained from tracking constitute the trajectories' number of positions (durations) $k_1, k_2, ...$ and estimated diffusion coefficients $D_1^{(e)}, D_2^{(e)}, ...$ (the superscript '(e)' is to distinguish the estimated diffusion coefficients of single trajectories from the parameters in the model to be estimated later), computed by

$$D_i^{(e)} = \frac{1}{4\Delta t(k_i - 1)} \sum_{n=1}^{k_i - 1} \left(\Delta x_{i,n}^2 + \Delta y_{i,n}^2 \right), \tag{1}$$

where $\Delta x_{i,n}$ and $\Delta y_{i,n}$ represent the n:th observed displacement in trajectory i. Figure 1 shows the experimental setup. By discretizing the estimated diffusion coefficients we can obtain a joint two-dimensional histogram of trajectory durations and diffusion coefficients. Of relevance to parameter estimation, we can also simulate corresponding data sets in a straightforward manner and generate simulated histograms.
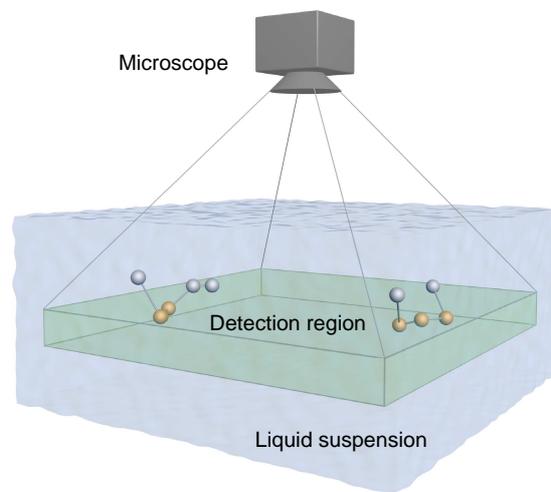
Figure 1: (Color online) A depiction of the experimental setup (not to scale). Particles inside the cuboid-shaped detection region are detected and tracked in two dimensions (x and y) using a confocal laser scanning microscope. The x and y (lateral) dimensions of the detection region, a_x and a_y, are known. The z (axial) dimension a_z needs to be estimated from the particle tracking data, and will depend on the choice of particle and other factors. The illustrated particles move inside (yellow) and outside (grey) of the detection region, yielding observed particle trajectories with two and three positions, respectively.

2.2 Parameter estimation

The principle of classical Bayesian (posterior) inference is as follows. For a data set $\mathcal{D} \in \mathcal{S}$ and a model with parameter vector $\boldsymbol{\theta}$, the posterior distribution $f(\boldsymbol{\theta}|\mathcal{D})$ for $\boldsymbol{\theta}$ is

$$f(\boldsymbol{\theta}|\mathcal{D}) \propto P(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{2}$$

with likelihood $P(\mathcal{D}|\boldsymbol{\theta})$ and prior distribution $\pi(\boldsymbol{\theta})$. The goal is to compute this posterior distribution, estimate the parameter $\boldsymbol{\theta}$ as a posterior mean (average of $\boldsymbol{\theta}$ over its distribution) and use percentiles of the distribution to obtain credible intervals i.e. error bounds. This is usually performed using some Monte Carlo method.

When the likelihood is either intractable or computationally expensive to compute, approximate Bayesian computation (ABC) is a useful simulation-based method of sampling from an approximate posterior distribution (Beaumont et al., 2002; Pritchard et al., 1999; Tavaré et al., 1997). If it is straightforward to generate simulated data from the model, sampling from the posterior can be performed in an accept-reject fashion by sampling a random $\boldsymbol{\theta}$ from $\pi(\boldsymbol{\theta})$, simulating a data set $\mathcal{D}'$ from the model for that parameter, and accepting $\boldsymbol{\theta}$ as a sample from the posterior if $\mathcal{D}' = \mathcal{D}$. It can be shown that this leads to sampling from the exact posterior. Because $P(\mathcal{D}' = \mathcal{D}|\boldsymbol{\theta})$ is extremely small in any interesting case, we relax the condition to accepting $\boldsymbol{\theta}$ as a sample from the (approximate) posterior if $\rho(\mathcal{D}, \mathcal{D}') \leq \epsilon$, where ρ is a suitable measure of discrepancy (a 'distance', but not necessarily a proper metric) and ϵ is a tolerance threshold. For $\epsilon = 0$ the exact posterior is obtained, whereas for larger ϵ, inference becomes more computationally feasible because more parameter candidates are accepted as samples from the approximate posterior; the choice of ϵ hence becomes a trade-off. Suggested sampling schemes for ABC include simple rejection, which is typically very slow, as well as Markov chain Monte Carlo ABC (Mar-

joram et al., 2003), partial rejection control ABC (Sisson et al., 2007, 2009), sequential Monte Carlo ABC (Toni et al., 2009), and population Monte Carlo ABC or PMC-ABC (Beaumont et al., 2009), the latter which we shall use herein. Population Monte Carlo approximate Bayesian computation (PMC-ABC) (Beaumont et al., 2009) is based on the population Monte Carlo method for standard Bayesian methods introduced by Cappé et al. (2004). The idea is to sequentially improve the approximation of the posterior distribution by introducing a decreasing sequence of tolerance thresholds ϵ_t, $1 \leq t \leq T$, and iteratively in each step performing a weighted resampling and a Gaussian-distributed perturbation of the parameter population in search of improvement in terms of ρ. It can be understood as a special case of population-based evolutionary optimization. For a population of N M-dimensional parameter vectors (M being twice the number of components/diffusion coefficients in our case), where $\boldsymbol{\theta}_n^{(t)}$ is the n:th parameter at iteration t, and $\theta_{n,m}^{(t)}$ is its m:th component, the algorithm looks like this (taken from (Beaumont et al., 2009) and somewhat adapted):

1. At iteration $t = 1$,

   (a) For $1 \leq n \leq N$, generate $\boldsymbol{\theta}_n^{(1)} \sim \pi(\boldsymbol{\theta})$ for each n, and simulate realizations $\mathcal{D}' \sim P(\mathcal{D}'|\boldsymbol{\theta}_n^{(1)})$ until $\rho(\mathcal{D}, \mathcal{D}') \leq \epsilon_1$.

   (b) For $1 \leq n \leq N$, let the weights (for weighted resampling) be $w_n^{(1)} = 1/N$.

   (c) For $1 \leq m \leq M$, let $\tau_m^{(1)}$ be $\sqrt{2}$ times the empirical standard deviation of the $\theta_{n,m}^{(1)}$.

2. At iterations $2 \leq t \leq T$,

   (a) For $1 \leq n \leq N$ and $1 \leq m \leq M$, randomly sample $\boldsymbol{\theta}_n^{\star}$ from the $\boldsymbol{\theta}_j^{(t-1)}$ with probabilities $w_j^{(t-1)}$ (weighted resampling), generate $\theta_{n,m}^{(t)} \sim \mathcal{N}\left(\theta_{n,m}^{\star}, \tau_m^{(t-1)}\right)$ and simulate realizations $\mathcal{D}' \sim P(\mathcal{D}'|\boldsymbol{\theta}_n^{(t)})$ until $\rho(\mathcal{D}, \mathcal{D}') \leq \epsilon_t$.

(b) For $1 \leq n \leq N$, let $w_n^{(t)} \propto \pi(\boldsymbol{\theta}_n^{(t)}) / \sum_{n'=1}^{N} w_{n'}^{(t-1)} \sum_{m=1}^{M} \frac{1}{\tau_m^{(t-1)}} \phi\left(\frac{1}{\tau_m^{(t-1)}}\left(\theta_{n,m}^{(t)} - \theta_{n',m}^{(t-1)}\right)\right)$,

where ϕ is the standard normal density.

(c) For $1 \leq m \leq M$, let $\tau_m^{(t)}$ be $\sqrt{2}$ times the empirical standard deviation of the $\theta_{n,m}^{(t)}$.

Moving to the problem at hand, we implement PMC-ABC in the following manner. We discretize the diffusion coefficients such that a data set can be thought of as a histogram. We define the discrepancy between a real data histogram $H_{\exp}(k,l)$ and a simulated data histogram $H_{\sim}(k,l)$ (k being the trajectory duration and l being the index of the bin for the discretized diffusion coefficients) as

$$\rho\left(H_{\exp}, H_{\sim}\right) = \sum_{k=k_{\min}}^{k_{\max}} \sum_{l=1}^{L} \left(H_{\exp}^{\text{cum}}(k,l) - H_{\sim}^{\text{cum}}(k,l)\right)^2 \tag{3}$$

where

$$H_{\exp}^{\text{cum}}(k,l) = \sum_{k'=k_{\min}}^{k} \sum_{l'=1}^{l} H_{\exp}(k',l') \tag{4}$$

and

$$H_{\sim}^{\text{cum}}(k,l) = \sum_{k'=k_{\min}}^{k} \sum_{l'=1}^{l} H_{\sim}(k',l') \tag{5}$$

are the *cumulative* histograms. We let $\gamma = \log_{10}(\epsilon)$ and decrease γ linearly in steps of $\Delta\gamma$, assuring in every iteration that $\log_{10}(\rho) \leq \gamma$ for the entire population. It is recognized that the choice of γ (ϵ) is arbitrary; we introduce some adaptivity by letting $\gamma_1 = \infty$ i.e. we start with generating a sample from the prior, then let γ_2 be the median of the $\log_{10}(\rho)$ values for $t = 1$, and after that decrease in steps of $\Delta\gamma = 0.01$. Further, we iterate until no further decrease of γ is possible without exceeding a maximum allowed number of simulation trials (in this work set to 500 times the number of posterior samples to be generated). The prior distribution is a uniform distribution over the parameters D_i, $a_{z,i}$, and $\log_{10}(c_i)$, for all

$1 \leq i \leq I$ (I being the number of components; internally, the software works with $\log_{10}(c)$ instead of c for numerical reasons). In our simulations, we let $A_x = A_y = 80$ μm and $A_z = 20$ μm.

2.3 Implementation details

Two implementations of the PMC-ABC algorithm are provided[*]. First, we provide a CPU implementation written in Julia (Bezanson et al., 2017), in which all loops over the N samples involving simulations and calculation of the distance ρ in Equation (3) are parallelized (but each simulation and distance calculation is executed serially). Second, we provide an optimized C++ and CUDA (NVIDIA Corporation, 2017) implementation that can off-load the simulations and calculations of the distance ρ to multiple graphics processing units (GPUs). The simulations are trivially parallelizable because simulated particles do not interact with each other; however, synchronization (e.g., via atomic operations) is required when simultaneously recording results in the histogram H_{sim}. Evaluation of the distance on the GPU avoids transferring H_{sim} from fast GPU memory to system memory. Both the cumulative $H_{\text{sim}}^{\text{cum}}$ histogram (Equation (5)) and the final distance ρ (Equation (3)) can be computed efficiently using parallel primitive operations (Hillis and Steele, 1986), that is, parallel reductions and prefix sums (Blelloch, 1990). Efficient implementations for both operations exist on GPUs. In particular, we take advantage of performing prefix sums with independent CUDA warps (Billeter et al., 2009), which eliminates the need for explicit synchronization inside the reduction and prefix sum operations. We ensure that multiple simulations and distance computations can run asynchronously and in parallel on all available GPUs via concurrent kernel execution (NVIDIA Corporation, 2017), to accommodate for parameter choices where the footprint of each simulation in isolation is relatively small and thus insufficient to fully utilize even modest GPUs.

---

[*]Source code for both is available at `https://github.com/roding/part_conc_spt_abc`.

2.4 Experiments

We use two species of fluorescent beads suspended in water. First, we use yellow/green fluorescent microspheres with actual size 0.175 ± 0.005 μm and concentration $\sim 3 \cdot 10^9$ part/ml (PS-Speck P7220, lot 1398540, Invitrogen, ThermoFisher Scientific, Waltham, MA, US). Second, we use blue/green/orange/dark red fluorescent microspheres with actual size 0.49 ± 0.015 μm and concentration $\sim 1.6 \cdot 10^9$ part/ml (TetraSpeck T7281, lot 1724855, Invitrogen, ThermoFisher Scientific, Waltham, MA, US). Third, we create a mixture of the two by mixing 3 μl solution of 0.175 μm beads with 6 μl solution of 0.49 μm beads. The resulting concentrations are intended to be $\sim 10^9$ part/ml and $\sim 0.8 \cdot 10^9$ part/ml, respectively. The solutions are sonicated for 20 min before use, and 7 μl of the solutions are placed in Secure-Seal imaging spacers (SigmaAldrich, Merck KGaA, Darmstadt, Germany) with 120 μm thickness. The beads are originally suspended in distilled water with 2 mM sodium azide and dilution is performed using distilled water. To study immobilized particles, the original suspensions are mixed with alginate gels (1 w/w % alginate acid sodium (Sigma-Aldrich, Steinheim, Germany), 0.88 w/w % NaCl, 0.555 w/w % GDL (D-(+)-Gluconic acid-δ-lactone (Sigma-Aldrich, Steinheim, Germany)), and 0.156 w/w % CaCO$_3$ (Acros Organics, Geel, Belgium)). These samples are placed in Secure-Seal imaging spacers as described above and allowed to rest over night until the particles are immobilized.

Particle tracking experiments are performed at ambient temperature using a Leica TCS SP5 II AOBS confocal laser scanning microscope (CLSM) with a DMI6000 inverted microscope (Leica Microsystems, Wetzlar, Germany). A 65 mW (average power) 488 nm Argon laser is used for excitation set at 20 % of full power with the AOTF set at 8 % (yielding an effective average power of ~ 1 mW). Imaging is performed with a photomultiplier tube detector (PMT, Hamamatsu Photonics, Hamamatsu, Japan), set to detect emission in the

range 496-620 nm with a gain voltage of 800 V. A Leica HC PL APO 63x/1.20 CS objective is used together with a zoom factor of 4, yielding a 61.5 μm x 61.5 μm field of view. The pinhole size is 1 Airy unit (111.5 μm for this objective). The resolution is 512 x 512 pixels with each pixel being 0.12 μm x 0.12 μm. A resonance scanner with 8000 Hz scanning rate is used, providing $\sim$ 14.7 fps ($\Delta t = 0.068$ s). For each data set, 25 videos of 1000 frames each are acquired at different locations in the sample; with 8-bit color depth, the data set size is 6.6 GB. The acquisition of a large number of independent videos at different locations minimizes the effect of local concentration fluctuations on the measurements. Measurements are performed at varying depths in the sample and at least 20 μm from the cover glass slip to avoid any boundary effects. Particle tracking is performed using TrackPy 0.3.2 (Tra). We use a lower threshold of 2000 for minimal mass (intensity sum) of each particle to exclude dim and diffuse particles. Only particle trajectories with 3 or more positions are kept. In Fig. 2, some trajectories are shown superimposed on a frame from one of the videos of the 0.175 μm beads. To assess localization error, we perform a bootstrapping-based nonparametric statistical analysis based on several particles, finding that localization error is $\sim$ 1.5 nm in one direction and $\sim$ 7 nm in the other (reflecting the anisotropy of acquisition due to the scanning procedure in CLSM). Consequently, localization error constitutes a very small fraction of the displacement of a particle between consecutive frames and is hence negligible. For the immobilized particles, the experimental setup is identical except a z-stack of 637 frames covering 80 μm in the axial direction is acquired.

We compute posterior means and credible intervals (using the 2.5 % and 97.5 % percentiles) of the D, a_z, c parameters. Additionally, we estimate the (hydrodynamic) size of the
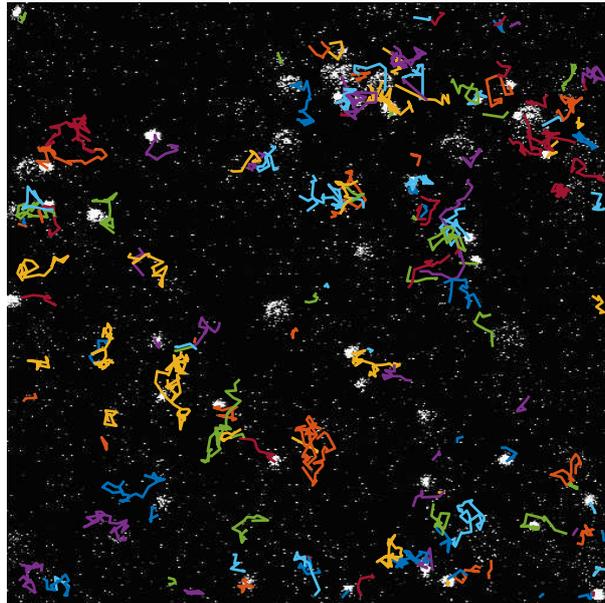
Figure 2: (Color online) Some trajectories are shown superimposed on a frame from one of the videos of the 175 nm beads. The field of view is 61.5 μm x 61.5 μm.

particle via the Stokes-Einstein relation: the diameter d can be written

$$d = \frac{k_B T}{3\pi\eta D}, \tag{6}$$

where k_B is Boltzmann's constant, $T = 293.15$ K is the (ambient) temperature, and $\eta = 8.9 \cdot 10^{-4}$ Pa $\cdot$ s is the viscosity of water at that temperature (Kestin et al., 1978). Converting all the samples from the posterior distribution for D to diameters yields a sample from the posterior distribution of d, for which we also compute posterior means and credible intervals. The results for the 0.175 μm beads are shown in Tab. 1 and the results for the 0.49 μm beads are shown in Tab. 2. Further, the full posterior distribution for the 0.175 μm case is shown in Fig. 3. As can be seen, the results are in good agreement with manufacturer specifications for concentration, and in very good agreement for size. Notably, the axial thickness of the detection region a_z is larger for the smaller particles; this is surprising because assuming a similar fluorescent dye, one would expect larger particles

13

Parameter	Posterior mean	Credible interval
D (μm/s^2)	2.65	$[2.63, 2.66]$
a_z (μm)	2.96	$[2.86, 3.07]$
c (part/ml)	$2.64 \cdot 10^9$	$[2.58 \cdot 10^9, 2.70 \cdot 10^9]$
Diameter (μm)	0.186	$[0.184, 0.187]$

Table 1: Parameter estimates for the 0.175 μm beads.

Parameter	Posterior mean	Credible interval
D (μm/s^2)	1.01	$[1.00, 1.02]$
a_z (μm)	2.53	$[2.43, 2.64]$
c (part/ml)	$1.60 \cdot 10^9$	$[1.55 \cdot 10^9, 1.66 \cdot 10^9]$
d (μm)	0.484	$[0.480, 0.489]$

Table 2: Parameter estimates for the 0.49 μm beads.

to be more strongly fluorescent (proportional to either volume or surface area) and thus be more easily detected away from the focal plane. However, the particles have different fluorescent labels, and the mass (intensity sum) of the particle positions as quantified by TrackPy is 7472 for the 0.175 μm particles and 6250 for the 0.49 μm. The ratio of these values is 1.20. The corresponding ratio of posterior mean estimates of the a_z parameter is 1.17. The fact that these two ratios are so similar suggests that a_z is strongly dependent on brightness.

In an attempt to validate the values of a_z in a more controlled setting, we also prepare samples of the monodisperse particles and let them rest over night until the particles are immobilized. Using the same experimental setting, we acquire z-stacks (with axial resolution 0.13 μm) and perform tracking of immobilized particles. Unfortunately, we find that these experiments are not directly comparable to the real tracking experiments. Here, the average a_z as estimated from several trajectories is ~ 3 μm for the 0.175 μm particles and ~ 4 μm for the 0.49 μm particles; the values are larger, and their order is inverted. The mass (intensity sum) of the particle positions is on average 8 % higher for the immobilized beads, explaining part of this effect. Further, we also estimate the point spread function (PSF) in the z direction, obtaining results between 1.8 μm and 4.0 μm for the 0.175 μm

14

particles and between 2.2 μm and 4.3 μm for the 0.49 μm particles, depending on the location in the sample. This has little impact on tracking because the a_z as estimated from immobilized beads were only weakly (negatively) correlated to the average z coordinates of the particles. However, we did find that the individual a_z values are broadly distributed and that these are correlated with the mean intensity of that particle ($\rho \approx 0.6$). This suggests that there is a distribution in fluorescence intensity and that a_z is really a distribution, even for a monodisperse sample.

Because the experiments on the monodisperse samples indicate that the axial thickness parameter varies significantly between the two species of particles, we analyze the 0.175 μm-0.49 μm bidisperse mixture solution allowing two different axial thickness parameters $a_{z,1}$ and $a_{z,2}$. The results for the mixture are shown in Tab. 3. Neglecting any uncertainty

Parameter	Posterior mean	Credible interval
D_1 (μm/s^2)	2.73	$[2.66, 2.80]$
$a_{z,1}$ (μm)	3.13	$[2.77, 3.47]$
c_1 (part/ml)	$7.49 \cdot 10^8$	$[6.89 \cdot 10^8, 8.14 \cdot 10^8]$
d_1 (μm)	0.180	$[0.175, 0.184]$
D_2 (μm/s^2)	1.05	$[1.02, 1.09]$
$a_{z,2}$ (μm)	2.78	$[2.44, 3.16]$
c_2 (part/ml)	$9.72 \cdot 10^8$	$[9.04 \cdot 10^8, 10.40 \cdot 10^8]$
d_2 (μm)	0.466	$[0.450, 0.482]$

Table 3: Parameter estimates for the 0.175 μm-0.49 μm bidisperse mixture.

in the dilution step, one would expect concentrations of $\sim 8.8 \cdot 10^8$ part/ml and $\sim 1.07 \cdot 10^9$ part/ml, hence the agreement between the concentrations estimated in the monodisperse and bidisperse cases are quite consistent with each other. The same goes for the other parameters: Estimated diffusion coefficients (and hence diameters) and axial thicknesses are all corresponding rather well to the values estimated from the monodisperse cases. We note however that the credible intervals broaden substantially when moving from one to two components, and we deem it unlikely that the model can successfully be extended to
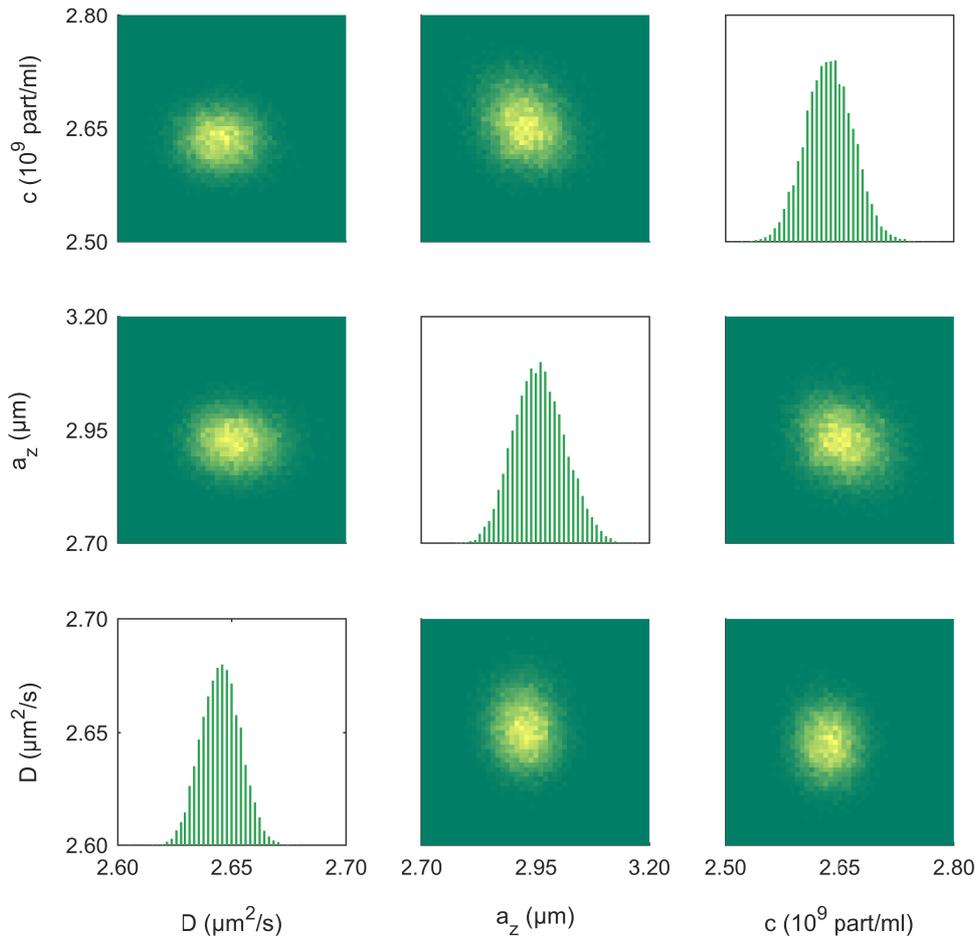
Figure 3: (Color online, double column figure) The posterior distribution of the parameters for the 0.175 μm beads. The diagonal subfigures show the marginal (one-dimensional) posterior distributions of the three individual parameters. The off-diagonal subfigures show the joint (two-dimensional) posterior distributions of each pair of two parameters.

16

three components and above at this point.

2.5 Performance benchmarks

We perform a couple of performance benchmarks to compare the CPU and GPU implementations, and in the GPU case also to assess the scaling with respect to the number of samples from the posterior and the number of GPUs used. The number of samples from the posterior is a trade-off; the larger the number, the more accurately can posterior means and credible intervals be estimated, but with larger computational cost. The benchmarks are all performed using the experimental data for the 175 nm beads.

We use two different setups for the benchmarks. The first, utilized for both CPU and GPU benchmarks, has dual Intel Xeon E5-2699 v4 CPUs (2×22 cores for a total of 88 hardware threads via hyper-threading) and two NVIDIA TITAN Xp GPUs. The second, utilized only for GPU benchmarks, is a compute cluster node with dual Intel Xeon E5-2690 CPUs and two or four NVIDIA Tesla K80 dual GK210-GPU accelerator cards. The latter is of interest primarily to assess scaling on many GPUs. Note that the GPU implementation uses very little CPU resources, as most of the time is spent waiting for GPU jobs to finish.

First, using $N = 4000$ samples, we run the CPU implementation using 88 parallel processes, the GPU implementation using 1 and 2 TITAN Xp GPUs, and using 1, 2, 4, and 8 GK210 GPUs. Run times are shown in Fig. 4.

Second, we run the GPU implementation on sample sizes between $N = 125$ and $N = 16000$, as shown in Fig. 5. As can be seen, the GPU implementation scales almost perfectly linearly with sample size.

The GPU implementation relies heavily on both integer operations and on transcendental ('special') function evaluation, for random number generation and transforming random
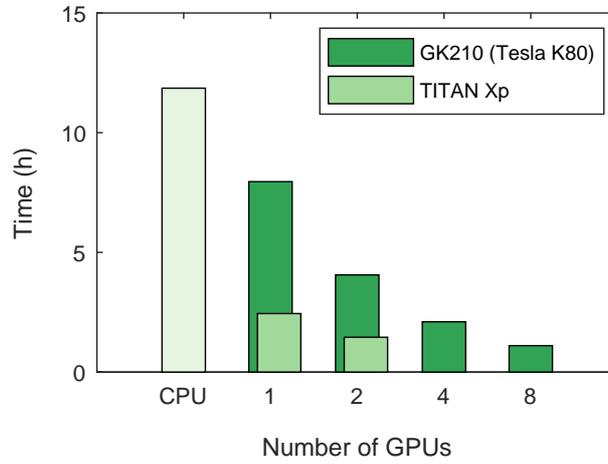
Figure 4: (Color online) Scaling of the run-time as a function of the number of GPUs and in comparison to the 88-thread dual-CPU reference setup.
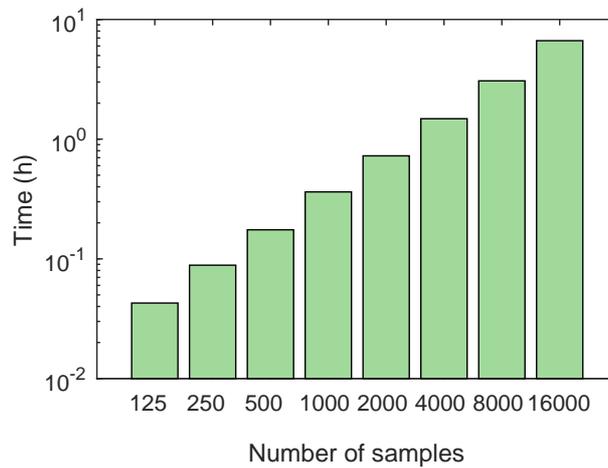

Figure 5: (Color online) Run-time scaling as a function of sample count for the GPU implementation. Notice the log scale on the y axis: the run-time scales linearly with sample count. Run-times were measured on the dual TITAN Xp GPU setup.

number distributions, respectively. When comparing a single TITAN Xp to a single GK210 GPU, the TITAN Xp performs slightly better than one would expect based on a simple extrapolation from peak single-precision floating-point performance. We believe this can be attributed to the increased throughput of special functions in the newer architecture of the TITAN Xp, and to differences in the handling of integer computations.

3 Conclusion

We have implemented a massively parallel population Monte Carlo approximate Bayesian computation (PMC-ABC) method for estimating diffusion coefficients, sizes and concentrations of diffusing nanoparticles in liquid suspension using a confocal laser scanning microscope (CLSM) and particle tracking. We have presented freely available CPU and GPU versions of the analysis software. Together with the commercial availability of CLSMs, this facilitates implementation and practical use of the method. We have applied the method to characterize mono- and bidisperse samples of fluorescent polystyrene beads and demonstrate good agreement with manufacturer-provided sizes and concentrations. We further demonstrate excellent scaling behavior for the algorithm on multiple GPUs. Interesting further work would for example be optimization of the thresholding schedule with respect to computational efficiency, and to take into account the fact the particles very close to the lateral dimension borders are not properly tracked; hence the effective field of view is actually slightly smaller than the theoretical one. Moreover, the results suggest that the axial size of the detection region, a_z, should be modeled as a distribution instead of a single value as done herein.

Acknowledgements

References

Trackpy 0.3.2. DOI:10.5281/zenodo.60550. https://zenodo.org/record/60550/files/trackpy-0.3.2.zip.

A. Banerjee, J. Qi, R. Gogoi, J. Wong, and S. Mitragotri. Role of nanoparticle size, shape and surface chemistry in oral drug delivery. *Journal of Controlled Release*, 238:176–185, 2016.

M. Beaumont, W. Zhang, and D. Balding. Approximate bayesian computation in population genetics. *Genetics*, 162:2025–2035, 2002.

M. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. Robert. Adaptive approximate bayesian computation. *Biometrika*, 96:983–990, 2009.

H. Berg. *Random walks in biology*. Princeton University Press, 1993.

J. Bezanson, A. Edelman, S. Karpinski, and V. Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59:65–98, 2017.

M. Billeter, O. Olsson, and U. Assarsson. Efficient stream compaction on wide simd many-core architectures. In *Proceedings of the Conference on High Performance Graphics 2009*, pages 159–166, 2009. ISBN 978-1-60558-603-8. doi: 10.1145/1572769.1572795. URL `http://doi.acm.org/10.1145/1572769.1572795`.

G. E. Blelloch. Prefix sums and their applications. Technical Report CMU-CS-90-190, CMU School of Computer Science, 1990. URL `https://www.cs.cmu.edu/~guyb/pubs.html`.

K. Braeckmans, K. Buyens, W. Bouquet, C. Vervaet, P. Joye, F. D. Vos, L. Plawinski, L. Doeuvre, E. Angles-Cano, N. Sanders, J. Demeester, and S. D. Smedt. Sizing nanomatter in biological fluids by fluorescence single particle tracking. *Nano letters*, 10: 4435–4442, 2010.

O. Cappé, A. Guillin, J.-M. Marin, and C. Robert. Population monte carlo. *Journal of Computational and Graphical Statistics*, 13:907–929, 2004.

B. Carr and M. Wright. Nanoparticle tracking analysis. *Innovations in Pharmaceutical Technology*, 26:38–40, 2008.

G. Chironi, C. Boulanger, A. Simon, F. Dignat-George, J.-M. Freyssinet, and A. Tedgui. Endothelial microparticles in diseases. *Cell and tissue research*, 335:143–151, 2009.

H. Deschout, K. Raemdonck, S. Stremersch, P. Maoddi, G. Mernier, P. Renaud, S. Jiguet, A. Hendrix, M. Bracke, R. van den Broecke, M. Röding, M. Rudemo, J. Demeester, S. C. de Smedt, F. Strubbe, K. Neyts, and K. Braeckmans. On-chip light sheet illumination enables diagnostic size and concentration measurements of membrane vesicles in biofluids. *Nanoscale*, 6:1741–1747, 2014.

L. Doeuvre, L. Plawinski, F. Toti, and E. Anglés-Cano. Cell-derived microparticles: a new challenge in neuroscience. *Journal of neurochemistry*, 110:457–468, 2009.

S. Du, K. Kendall, S. Morris, and C. Sweet. Measuring number-concentrations of nanoparticles and viruses in liquids on-line. *Journal of chemical technology and biotechnology*, 85:1223–1228, 2010.

V. Filipe, R. Poole, M. Kutscher, K. Forier, K. Braeckmans, and W. Jiskoot. Fluorescence single particle tracking for the characterization of submicron protein aggregates in biological fluids and complex formulations. *Pharmaceutical research*, 28:1112–1120, 2011.

V. Filipe, R. Poole, O. Oladunjoye, K. Braeckmans, and W. Jiskoot. Detection and characterization of subvisible aggregates of monoclonal IgG in serum. *Pharmaceutical research*, 29:2202–2212, 2012.

K. Forier, A.-S. Messiaen, K. Raemdonck, H. Nelis, S. D. Smedt, J. Demeester, T. Coenye, and K. Braeckmans. Probing the size limit for nanomedicine penetration into burkholderia multivorans and pseudomonas aeruginosa biofilms. *Journal of Controlled Release*, 195:21–28, 2014.

W. D. Hillis and G. L. Steele, Jr. Data parallel algorithms. *Communications of the ACM*, 29(12):1170–1183, 1986. ISSN 0001-0782. doi: 10.1145/7902.7903. URL `http://doi.acm.org/10.1145/7902.7903`.

K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. Schmid, and G. Danuser. Robust single-particle tracking in live-cell time-lapse sequences. *Nature methods*, 5:695–702, 2008.

J. Kestin, M. Sokolov, and W. Wakeham. Viscosity of liquid water in the range -8 c to 150 c. *Journal of Physical and Chemical Reference Data*, 7:941–948, 1978.

S. Khare and N. Dokholyan. Molecular mechanisms of polypeptide aggregation in human diseases. *Current Protein and Peptide Science*, 8:573–579, 2007.

S. Kumar, A. Anselmo, A. Banerjee, M. Zakrewsky, and S. Mitragotri. Shape and size-dependent immune response to antigen-carrying nanoparticles. *Journal of Controlled Release*, 220:141–148, 2015.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100:15324–15328, 2003.

I. Montes-Burgos, D. Walczyk, P. Hole, J. Smith, I. Lynch, and K. Dawson. Characterisation of nanoparticle size and state prior to nanotoxicological studies. *Journal of Nanoparticle Research*, 12:47–53, 2010.

B. Naeye, H. Deschout, V. Caveliers, B. Descamps, K. Braeckmans, C. Vanhove, J. Demeester, T. Lahoutte, S. D. Smedt, and K. Raemdonck. In vivo disassembly of iv administered sirna matrix nanoparticles at the renal filtration barrier. *Biomaterials*, 34: 2350–2358, 2013.

A. Nel, L. Mädler, D. Velegol, T. Xia, E. Hoek, P. Somasundaran, F. Klaessig, V. Castranova, and M. Thompson. Understanding biophysicochemical interactions at the nano–bio interface. *Nature materials*, 8:543–557, 2009.

S. Nune, P. Gunda, P. Thallapally, Y.-Y.Lin, F. Laird, and C. Berkland. Nanoparticles for biomedical imaging. *Expert opinion on drug delivery*, 6:1175–1194, 2009.

NVIDIA Corporation. CUDA C programming guide, 2017. URL `http://docs.nvidia.com/cuda/cuda-c-programming-guide/`.

J. Pritchard, M. Seielstad, A. Perez-Lezaun, and M. Feldman. Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular Biology and Evolution*, 16:1791–1798, 1999.

K. Remaut, N. Sanders, B. D. Geest, K. Braeckmans, J. Demeester, and S. D. Smedt. Nucleic acid delivery: Where material sciences and bio-sciences meet. *Materials Science and Engineering: R: Reports*, 58:117–161, 2007.

M. Röding, H. Deschout, K. Braeckmans, and M. Rudemo. Measuring absolute number concentrations of nanoparticles using single-particle tracking. *Physical Review E*, 84(3): 031920, 2011.

M. Röding, H. Deschout, K. Braeckmans, and M. Rudemo. Measuring absolute nanoparticle number concentrations from particle count time series. *Journal of microscopy*, 251: 19–26, 2013a.

M. Röding, H. Deschout, K. Braeckmans, A. Särkkä, and M. Rudemo. Self-calibrated concentration measurements of polydisperse nanoparticles. *Journal of microscopy*, 252: 79–88, 2013b.

M. Röding, E. Zagato, K. Remaut, and K. Braeckmans. Approximate bayesian computation for estimating number concentrations of monodisperse nanoparticles in suspension by optical microscopy. *Physical Review E*, 93:063311, 2016.

S. Salatin, S. Dizaj, and A. Khosroushahi. Effect of the surface modification, size, and shape on cellular uptake of nanoparticles. *Cell biology international*, 39:881–890, 2015.

H. Saveyn, B. D. Baets, O. Thas, P. Hole, J. Smith, and P. V. D. Meeren. Accurate particle size distribution determination by nanoparticle tracking analysis based on 2-d brownian dynamics simulation. *Journal of colloid and interface science*, 352:593–600, 2010.

L. Shang, K. Nienhaus, and G. Nienhaus. Engineered nanoparticles interacting with cells: size matters. *Journal of nanobiotechnology*, 12:5, 2014.

S. Sisson, Y. Fan, and M. Tanaka. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104:1760–1765, 2007.

S. Sisson, Y. Fan, and M. Tanaka. Correction to "sequential monte carlo without likelihoods". *Proceedings of the National Academy of Sciences*, 106:16889, 2009.

S. Tavaré, D. Balding, R. Griffiths, and P. Donnelly. Inferring coalescence times from dna sequence data. *Genetics*, 145:505–518, 1997.

T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6:187–202, 2009.

R. Weissleder, M. Nahrendorf, and M. Pittet. Imaging macrophages with nanoparticles. *Nature materials*, 13:125–138, 2014.

S. Wilhelm, A. Tavares, Q. Dai, S. Ohta, J. Audet, H. Dvorak, and W. Chan. Analysis of nanoparticle delivery to tumours. *Nature Reviews Materials*, 1:16014, 2016.

J.-W. Yoo, D. J. D.J. Irvine, D. E. Discher, and S. Mitragotri. Bio-inspired, bioengineered and biomimetic drug delivery carriers. *Nature reviews Drug discovery*, 10, 2011.

E. Zagato, K. Forier, T. Martens, K. Neyts, J. Demeester, S. D. Smedt, K. Remaut, and K. Braeckmans. Single-particle tracking for studying nanomaterial dynamics: applications and fundamentals in drug delivery. *Nanomedicine*, 9:913–927, 2014.