Supplementary Information (SI) for Soft Matter.

Supplementary Information

Measuring collective diffusion coefficients by counting particles in boxes

A. Carter, E. K. R. Mackay, B. Sprinkle, A. Thorneywork, S. Marbach

April 16, 2025

Supplementary Discussion

1 Methodological details **2**
1.1 Experiments. 2
1.2 Simulations. 2
1.3 Calculation of dynamic structure factors. 2

2 Alternate method for obtaining $D(L)$ **3**
2.1 Two regimes of box sizes: small for individual dynamics and large for collective. 3
2.2 Phenomenological fit of number fluctuations 5

3 Supplementary theory **6**
3.1 Properties of timescale integrals 6
3.2 Wavenumber dependent $D(k)$ predictions 7
3.3 Static structure factor 7

4 Supplementary analysis **9**
4.1 Countoscope rescaling in the dilute regime 9
4.2 Countoscope plateau value 10
4.3 Effect of simulation box size on $D(L)$, $D(k)$ 12
4.4 $f(k,t)$ behaviour at short and long times 13
4.5 $f(k,t)$ short time divergence 14

1 Methodological details

1.1 Experiments.

Carboxylate-functionalised melamine formaldehyde particles (Microparticles GmbH) are dispersed in 20/80 v/v% ethanol-water mixtures, and suspensions are loaded into quartz glass flow cells. The particles' high mass density ($\rho = 1510\,\mathrm{kg\,m^{-3}}$) means they sediment rapidly to the base of the sample cell to form a monolayer, and further that the gravitational height is small ($\approx 0.02\,\sigma$) so out-of-plane fluctuations are negligible. Due to this close confinement, the particles interact hydrodynamically with the base of the flow cell, but interactions with the upper surface are not significant as this is separated by a height of $\approx 70\,\sigma$ from the plane of the particles. The flow cell is imaged with a custom-built inverted microscope, as in Ref. [8]. Microscope images provide a $390 \times 295\mathrm{\mu m}^2$ field of view, approximately 10^{-6} times smaller than the total sample cell area. The pixel size is $0.288\,\mathrm{\mu m}$ and timestep between frames $0.5\,\mathrm{s}$. We ignore any interactions with the flow cell side walls. Particle diameters are characterised by fitting hard-disk DFT results [9] to the experimental structure factor (see Fig. S2). The imaging stage tilt is adjusted iteratively until no systemic drift in particle positions is measured. Data sets are checked for residual drift in positions or image packing fraction over the course of the experiments.

1.2 Simulations.

Simulations were performed as described in Ref. [6]. The simulation box was of side length $1280\,\mathrm{\mu m}$ except where specified, and particle diameter $3.0\,\mathrm{\mu m}$. Particle trajectories are wrapped with periodic boundary conditions in x and y. Simulations were run twice, once with data being saved for analysis with a timestep of $0.5\,\mathrm{s}$ and simulated time 22 hours, and subsequently with a timestep of $64\,\mathrm{s}$ and simulated time of 710 hours. Data from the short runs were used to compute $D(k)$ for short length scales ($k > 1.8/\sigma$), and data from the long runs used to compute $D(k)$ for large length scales ($k < 1.8/\sigma$). As we compute $D(k)$ from inverting the first non-zero time point of $f(k,t)$, for small k we are probing $D(k, 0.5\,\mathrm{s})$, but for large k we are probing $D(k, 64\,\mathrm{s})$. This is the source of the small discontinuity visible on Fig. 6 at $L/\sigma = 3.5$. For the timescale integral, the short runs are used to calcuate $D(L)$ for $L < 1.3\sigma$, and the long runs for $L > 1.3\sigma$. For the phenomenological fit, the number fluctuations $\langle \Delta N^2(t) \rangle$ are combined, with data from the short runs used for $t < 1024\,\mathrm{s}$ and data from the long runs used for $t \geq 1024\,\mathrm{s}$. The fit is then performed on this combined dataset of number fluctuations.

1.3 Calculation of dynamic structure factors.

We give further details on how we compute $f(k,t)$. First, Eq. (8) is extended to 2-time points with non-constant N_p particles, as particle numbers in experiments are not conserved over time. We take the normalizing factor as the average over the 2 time points, $N_p = (N_p(0) + N_p(t))/2$, and the sum runs over all particles at each time frame. The dynamic structure factor is first computed via Eq. (8) over a 2D grid of k vectors as $F(k_x, k_y, t)$ where $(k_x, k_y) = (n2\pi/L_x, m2\pi/L_y)$, the size of the optical image is (L_x, L_y), and n, m are integers. To reduce computational expense, not all values of n and m are used for large k, where the points are located very close to each other in log-space. The wavevector magnitude is then calculated as $k = \sqrt{{k_x}^2 + {k_y}^2}$, and a radial average is performed by binning the values of k and F with bins spaced as above, and their final values are taken as the mean of their respective values in each bin.

2 Alternate method for obtaining $D(L)$

In the main text (section II A) we described how to obtain $D(L)$ from integrating the correlation function $C_N(t)$. This method produces more accurate results but requires lengthy datasets to properly resolve the correlation function at long time.

It is also possible to obtain $D(L)$ by making a fit to $\langle \Delta N^2(t) \rangle$, as we describe here. This method works on shorter length datasets, but requires knowledge of a model of the correlation function.

2.1 Two regimes of box sizes: small for individual dynamics and large for collective.

Theory highlights two limit regimes of box sizes. To understand how to probe a meaningful quantity from number fluctuations, we first examine the behaviour of $\langle \Delta N^2(t) \rangle$ in the light of theory. Predictions for such correlation functions can be obtained using stochastic density field theory (Dean-Kawasaki equations) [1, 5].

In the absence of interparticle interactions, some of us have shown ([6, 7]) that $\langle \Delta N^2(t) \rangle$ obeys the analytical law

$$\langle \Delta N^2(t) \rangle = 2\langle N \rangle \left(1 - \left[f\left(\frac{4D_{\mathrm{self}} t}{L^2} \right) \right]^2 \right), \tag{S1}$$

where f has the explicit expression

$$f\left(\tau = \frac{4D_{\mathrm{self}} t}{L^2} \right) = \sqrt{\frac{\tau}{\pi}} \left(e^{-1/\tau} - 1 \right) + \mathrm{erf}\left(\sqrt{1/\tau} \right) \tag{S2}$$

and in the absence of interactions we recall $D_{\mathrm{self}} = D_0$. Eq. (S1) perfectly agrees with experimental data in the dilute regime (see Fig. S3-a).

To account for steric repulsion between particles [6], Eq. (S1) can be modified, leading to a slightly more complicated expression for $\langle \Delta N^2(t) \rangle$,

$$\langle \Delta N^2(t) \rangle = 2\langle N \rangle \int \frac{k \mathrm{d}k}{(2\pi)^2} f_{\mathcal{V}}(k) S(k) - 2C_N(t) \tag{S3}$$

where

$$C_N(t) = \langle N \rangle \int \frac{k \mathrm{d}k}{(2\pi)^2} L^2 f_{\mathcal{V}}(k) S(k) e^{-D_{\mathrm{self}} k^2 t / S(k)} \tag{S4}$$

and $f_{\mathcal{V}}(k)$ is a characteristic non dimensional geometrical factor depending on the wavenumber amplitude only

$$f_{\mathcal{V}}(k) = \int \mathrm{d}\theta \left(\frac{\sin(k_x(\theta)L/2)}{k_x(\theta)L/2} \right)^2 \left(\frac{\sin(k_y(\theta)L/2)}{k_y(\theta)L/2} \right)^2 \tag{S5}$$

with $k_x(\theta) = k\cos\theta$ and $k_y(\theta) = k\sin\theta$; and $S(k)$ is the structure factor of the fluid. This modified expression is hard to use directly, as it involves the suspension's structure factor $S(k)$ for a given wavelength k.

Our goal here is to identify the limit laws that Eq. (S4) converges to in the limit of small and large boxes respectively. To do so, we first introduce the non-dimensional wavenumber $K = kL/2$ such that the geometrical factor becomes

$$f_{\mathcal{V}}(K) = \int \mathrm{d}\theta \left(\frac{\sin(K\cos\theta)}{K\cos\theta} \right)^2 \left(\frac{\sin(K\sin\theta)}{K\sin\theta} \right)^2 \tag{S6}$$

and the correlation function

$$C_N(t) = \langle N \rangle \int \frac{K \mathrm{d}K}{(\pi)^2} f_{\mathcal{V}}(K) S\left(\frac{2K}{L} \right) e^{-(4D_{\mathrm{self}} t/L^2) \times K^2 / S\left(\frac{2K}{L}\right)}. \tag{S7}$$

It is then useful to approximate the structure factor $S(k)$ in the following way. Typically, when $k \geq 1/\sigma$, $S(k) \simeq 1$ and when $k \leq 1/\sigma$, then $S(k) \simeq S(k=0)$. This means we can split the above integral in 2 distinct contributions

$$\begin{aligned} C_N(t) = & \langle N \rangle \int_{L/2\sigma}^{\infty} \frac{K \mathrm{d}K}{(\pi)^2} f_{\mathcal{V}}(K) e^{-\frac{4D_{\mathrm{self}}t}{L^2}K^2} + \\ & \langle N \rangle S(0) \int_0^{L/2\sigma} \frac{K \mathrm{d}K}{(\pi)^2} f_{\mathcal{V}}(K) e^{-\frac{4D_{\mathrm{self}}t}{S(0)L^2}K^2} . \end{aligned} \tag{S8}$$

It is then rather straightforward to look at the two limiting cases of small and large box. If $L \lesssim \sigma$, then $L/\sigma \lesssim 1$. Clearly, the 2nd integral in Eq. (S8) is negligible, and one can assume $\int_{L/2\sigma}^{\infty} \simeq \int_0^{\infty}$. We arrive at an expression corresponding to the 2D case without interactions, where the integral over K can be conducted, and

$$C_N(t) = \langle N \rangle \left[f\left(\frac{4D_{\mathrm{self}}t}{L^2} \right) \right]^2 , \tag{S9}$$

where f is defined in Eq. (S1). Using Eq. (4) of the main text, one directly arrives at Eq. (S1).

In contrast, when $L \gtrsim \sigma$, then $L/\sigma \gtrsim 1$. In that case, the 1st integral in Eq. (S8) becomes negligible and we can assume $\int_0^{L/2\sigma} \simeq \int_0^{\infty}$. We also arrive at an expression corresponding to the 2D case without interactions, but this time $D_{\mathrm{self}} \to D_{\mathrm{self}}/S(0)$ and the prefactor $S(0)$ is in front of the integration sign. This means we obtain

$$C_N(t) = \langle N \rangle S(0) \left[f\left(\frac{4D_{\mathrm{self}}t}{S(0)L^2} \right) \right]^2 . \tag{S10}$$

Since $D_{\mathrm{coll}} = D_{\mathrm{self}}/S(0)$, we can replace it in the expression above. In addition, because $C_N(0) = \mathrm{Var}(N)$, and $f(0) = 1$, one sees that in the limit of large boxes $L \gtrsim \sigma$

$$\mathrm{Var}(N) \simeq \langle N \rangle S(0) \tag{S11}$$

which is a well-established link in liquid matter [3, 4, 6]. For large boxes we arrive at

$$C_N(t) = \mathrm{Var}(N) \left[f\left(\frac{4D_{\mathrm{coll}}t}{L^2} \right) \right]^2 , \tag{S12}$$

and using Eq. (4), we obtain

$$\langle \Delta N^2(t) \rangle \simeq 2\,\mathrm{Var}(N) \left(1 - \left[f\left(\frac{4D_{\mathrm{coll}}t}{L^2} \right) \right]^2 \right) . \tag{S13}$$

Finally, one can also notice that for small boxes $L \lesssim \sigma$, based on Eq. (S9),

$$\mathrm{Var}(N) \simeq \langle N \rangle . \tag{S14}$$

Individual and collective regimes are fully apparent in simulations and experiments. To confirm the presence of these limiting regimes, we plot $\langle \Delta N^2(t) \rangle$ in the intermediate density system, rescaling time by L^2 and $\langle \Delta N^2(t) \rangle$ also by L^2, for simulations and experiments (Fig. S1-a and b). On each plot, we also present the limiting regimes Eq. (S1) (dotted) and Eq. (S13) (dashed). In the simulations, small and large box data indeed converge to the limiting regimes. In the experiments, the agreement with the collective regime is less clear. This is in part due to fewer statistics on the experimental data; but could also be due to collective hydrodynamic effects, that are not accounted for in the theory or in the simulations in Fig. S1-a. Nonetheless, curves converge to a limit regime at large length scales.

2.2 Phenomenological fit of number fluctuations

Workflow of the phenomenological fit method To investigate the box-size dependent relaxation speed, we thus seek a box-size dependent diffusion coefficient $D(L)$. Considering the similarity between Eq. (S1) and Eq. (S13), we introduce the *phenomenological fitting law* for any box size with the ansatz

$$\langle \Delta N^2(t) \rangle = 2\,\mathrm{Var}(N) \left(1 - \left[f\left(\frac{4D(L)t}{L^2} \right) \right]^2 \right) \tag{S15}$$

where $D(L)$ is a fitting parameter at every box size. Note that since $\mathrm{Var}(N) = \langle N \rangle$ for small boxes (Fig. S3), then Eq. (S15) indeed describes number fluctuations of small boxes provided $D(L \ll \sigma) = D_{\mathrm{self}}$. Eq. (S15) also converges to the large box regime, provided $D(L \gg \sigma) = D_{\mathrm{coll}}$.

Scale-dependent diffusion coefficient $D(L)$ from the phenomenological fit. Based on this phenomenological fitting law, one can obtain predictions for the scale-dependent diffusion coefficient $D(L)$ as shown in Fig. S1-c and d, again for simulations and experiments. For reference, we also add theory curves for $D(L)$, as established from the timescale integral method in the main text. We recall that the theory curves converge to D_{coll} for $L \gg \sigma$, and we use mostly this limit as a reference. For the intermediate density regime $\phi = 0.11$, the theory predicts $D_{\mathrm{coll}} = D_{\mathrm{self}}/S(k=0) \simeq 1.6 D_{\mathrm{self}}$ (dotted lines in Fig. S1-c and d).

In the intermediate density regime $\phi = 0.11$, $D(L)$, for simulations and experiments, interpolates between 2 limiting cases: $D(L)$ plateaus for small boxes, and then increases before reaching another larger plateau. For small boxes $L \ll \sigma$, we find $D(L) \simeq D_{\mathrm{self}}$, the box-dependent diffusion coefficient probes individual motion, as expected. In contrast, in larger boxes, group motion dominates and increases the effective diffusion coefficient $D(L)$. The hard sphere repulsion between particles accelerates group diffusion: a single particle's motion can influence that of its neighbours via repulsion, and of the subsequent neighbours.

The plateau reached in large boxes for both simulated and experimental data is much larger than the expected D_{coll}. This is likely an artefact of the phenomenological fitting. At short enough times, the number fluctuations are dominated by individual particles entering and exiting a box – curves all agree with Eq. (S1) (Fig. S1, dotted black line). The phenomenological fit uses a single diffusion coefficient to model both this early-time individual behaviour and long-time collective, resulting in a slightly larger $D(L)$ for large boxes than D_{coll}, given that the variance is kept constant. The phenomenological fit could thus be improved by fitting long-time evolution only, but for simplicity, we do not investigate this here. The phenomenological fit Eq. (S15) thus only provides qualitative insights into the emergence of collective phenomena at large scales.

The variations in $D(L)$ are less apparent in the dilute regime, $\phi = 0.02$, which is expected as collective effects do not significantly modify dynamics at this low packing fraction. Nonetheless, a slight increase of $D(L)$ is still observed, showing the sensitivity of the technique even to mild differences in collective versus individual properties.

Comparison between timescale integral and phenomenological fit. We now compare the phenomenological fitting approach with the timescale integral approach. Overall, both methods predict somewhat similar shapes of $D(L)$ and limiting regimes, which confirms their ability to sense collective effects at a given scale. The results from the phenomenological fit (Fig. S1-c and d) are more distant to the theory than those from the timescale integral. This confirms that the timescale integral approach is more accurate than the phenomenological fit, provided one has long enough data sets.

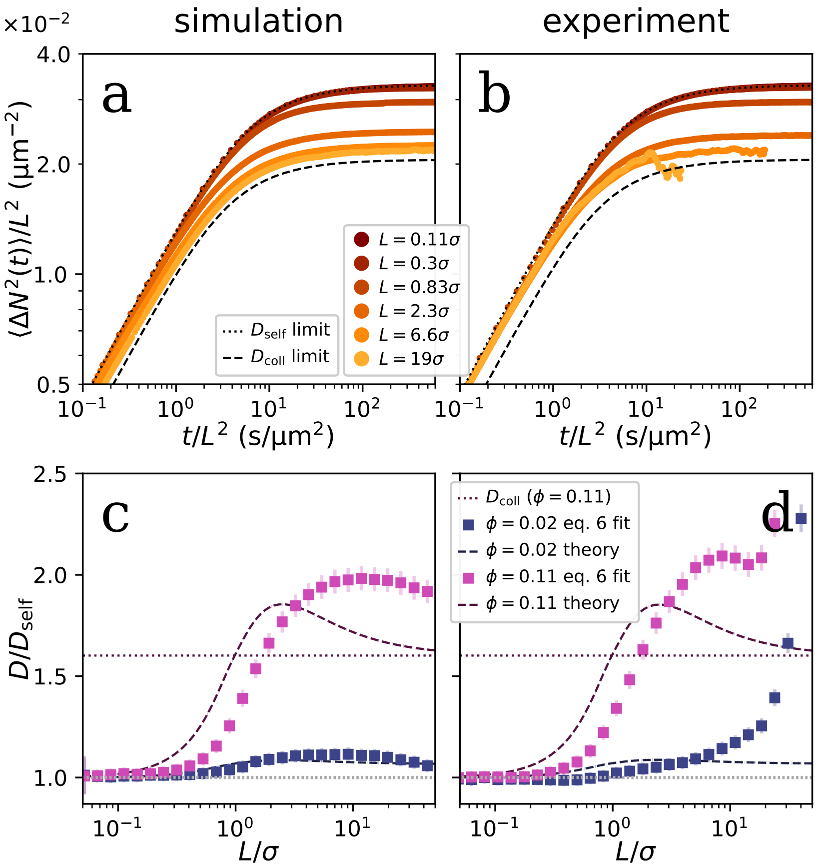

Figure S1: Phenomenological fitting of number fluctuations. (a-b) Rescaled $\langle \Delta N^2(t) \rangle$ versus rescaled time for (a) simulations and (b) experimental data in the intermediate density $\phi = 0.11$ regime. Box sizes go from small (dark red) to large (yellow). Dashed curves represent the limit regimes Eq. (S1) and Eq. (S13). (c-d) Obtained $D(L)$ from $\langle \Delta N^2(t) \rangle$ fits of Eq. (S15) for (c) simulations and (d) experimental data. Lines correspond to the theory Eq. (6). Blue corresponds to the dilute regime, and pink the intermediate density.

Although the timescale integral appears more robust, one method might be more convenient than the other in practical situations. Because the phenomenological fit investigates $\langle \Delta N^2(t) \rangle$, it focuses on early to intermediate time scales, requiring somewhat shorter datasets for fitting. However, it works under the assumption of a model. In contrast, the timescale integral does not require any model *a priori*, and therefore, could suitably be used to infer a relaxation timescale $T(L)$ of any underlying physical process. The downside is that it requires the long-time decay of the correlation functions to be highly resolved to obtain convergence of the integral. Overall, we find that using both approaches could be useful as the point where they differ can serve as an indicator for when the data is not sufficient to obtain accurate predictions.

3 Supplementary theory

3.1 Properties of timescale integrals

Given the expression of the timescale integral in Eq. (5), it is then straightforward, with Eq. (S9), to show that in the limit of small boxes,

$$T(L) = \alpha_T \frac{L^2}{4D_{\text{self}}} \text{ where } \alpha_T = \int_0^\infty f(\tau)^4 \mathrm{d}\tau. \tag{S16}$$

Similarly, in the limit of large boxes, one arrives at a similar expression

$$T(L) = \alpha_T \frac{L^2}{4D_{\text{coll}}} \tag{S17}$$

and hence we obtain the following Countoscope limits

$$D(L \to 0) = D_{\text{self}},\ D(L \to \infty) = D_{\text{coll}}. \tag{S18}$$

3.2 Wavenumber dependent $D(k)$ predictions

We here show that we expect $D(k) = D_{\text{self}}/S(k)$ within the theory we used to investigate hard sphere dynamics. The first step is to notice that the correlation function is obtained from integrating the density correlation function in Fourier space, as

$$\begin{aligned} C_N(t) &= \int_{[-\frac{L}{2},\frac{L}{2}]^2} \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r'} \langle (\rho(\boldsymbol{r},t) - \rho_0)(\rho(\boldsymbol{r'},0) - \rho_0) \rangle, \\ &= \rho_0 \int_{[-\frac{L}{2},\frac{L}{2}]^2} \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r'} \int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^2} e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r'})} F(k,t). \end{aligned}$$

Going back to cartesian coordinates, with $k_x(\theta) = k\cos\theta$ and $k_y(\theta) = k\sin\theta$, and integrating over the box gives

$$\begin{aligned} C_N(t) &= \int_{[-\frac{L}{2},\frac{L}{2}]^2} \mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y' \int \frac{k\mathrm{d}k\mathrm{d}\theta}{(2\pi)^2} e^{ik_x(\theta)(x-x')} e^{ik_y(\theta)(y-y')} F(k,t) \\ &= \rho_0 L^4 \int \frac{k\mathrm{d}k}{(\pi)^2} f_{\mathcal{V}}(k) F(k,t), \end{aligned}$$

where $f_{\mathcal{V}}(k)$ is defined in Eq. (S5). Given that $\langle N \rangle = \rho_0 L^2$, we can identify terms with Eq. (S4), and find

$$F(k,t) = S(k) e^{-D_{\text{self}} k^2 t / S(k)}. \tag{S19}$$

Note that expression is only valid within an approximation of *small* density fluctuations – a linearization of the so-called Dean Kawasaki equation [1, 5], which is the starting point of our theory. Other approaches also stress that this is a linear approximation [2]. It is then straightforward to identify from Eq. (8):

$$D(k,t) = \frac{D_{\text{self}}}{S(k)}. \tag{S20}$$

Note that within this approximation, $D(k,t)$ does not depend on time, and we can simply write $D(k)$.

3.3 Static structure factor

In Eq. (S4) and onwards we use an analytic expression for the structure factor of hard spheres in 2D, which is based on density field theory and is in remarkable agreement with our 2D-sedimented colloidal experiments, as was verified in a previous work [10] and can be seen in Fig. S2. We report it here for consistency:

$$S(k) = \frac{1}{1 - \rho c^{(2)}(k)} \tag{S21}$$

where ρ is the mean particle density and

$$\begin{aligned}c^{(2)}(k) =& \frac{\pi}{6(1-\phi)^3k^2}\Bigg[-\frac{5}{4}(1-\phi)^2k^2\sigma^2J_0(k\sigma/2)^2 \\ &+\Bigg(4\left((\phi-20)\phi+7\right) \\ &+\frac{5}{4}(1-\phi)^2k^2\sigma^2\Bigg)J_1(k\sigma/2)^2 \\ &+2(\phi-13)(1-\phi)k\sigma J_1(k\sigma/2)J_0(k\sigma/2)\Bigg]\end{aligned} \tag{S22}$$

where $J_i(x)$ are Bessel functions of the first kind. The limit of vanishing wavenumber in Eq. (S22) can be taken analytically

$$S(k=0) = \frac{(1-\phi)^3}{1+\phi} \tag{S23}$$

and is consistent with the result from the scaled particle theory equation of state.

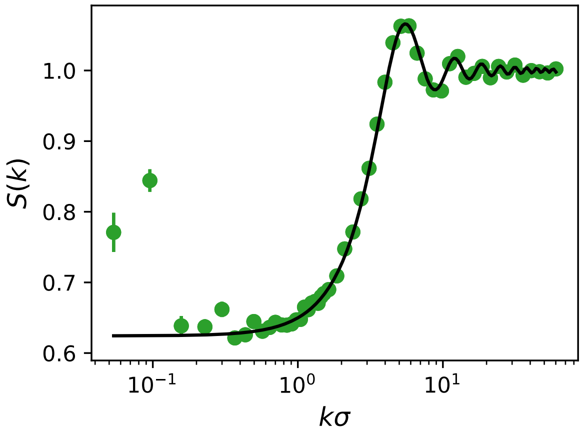

Figure S2: $S(k)$ for experimental data at packing fraction $\phi = 0.11$ (green). Black curve is a fit to Eq. S21 where the particle diameter σ is the only adjustable parameter. We find $\sigma = 3.0\,\mu\text{m}$.

4 Supplementary analysis

4.1 Countoscope rescaling in the dilute regime

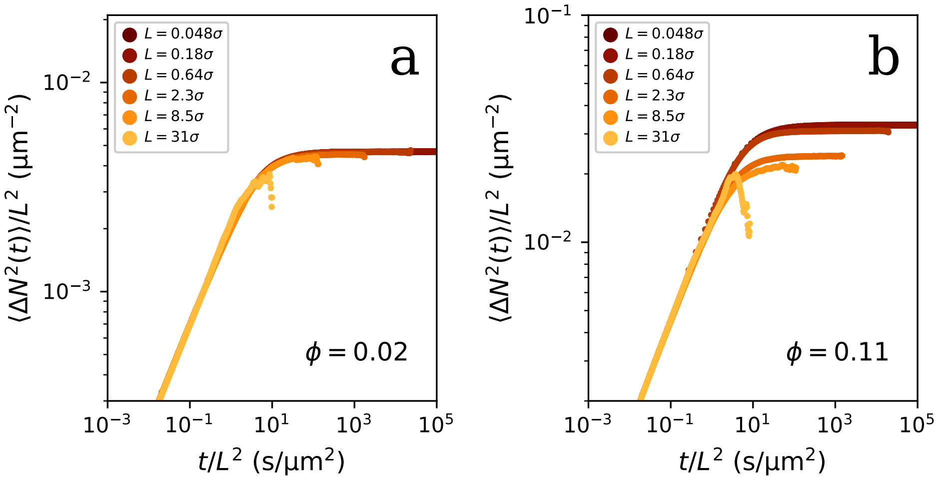

Figure S3: (a) at $\phi = 0.02$, the data rescales onto itself for all box sizes. The black line corresponds with the theory Eq. (S1) in the main text. Points correspond to experimental data. This rescaling also illustrates that the plateau value corresponds to two times the mean number of particles in a box in the dilute regime, mathematically $\mathrm{Var}(N) = \langle N \rangle$. (b) at $\phi = 0.11$, the data does not rescale onto itself, indicating that large boxes are probing a different effect to small boxes, namely collective effects.

4.2 Countoscope plateau value

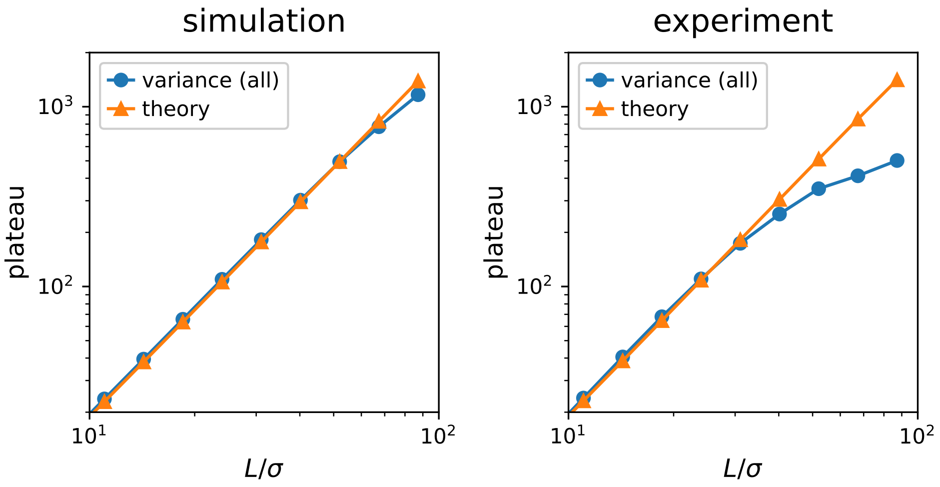

Figure S4: Plateau values against box size for simulations and experiments. The plateau is defined as $2\text{Var}(N)$. It is either measured by computing $\text{Var}(N)$ over all boxes and all times (blue) or calculated via the expression of $C_N(t=0) = \text{Var}(N)$ of Eq. (S4) (orange). Packing fraction is $\phi = 0.11$. The deviation of the variance from it's theoretical value for experiments shows that we do not have enough data to properly resolve the variance at large L, a problem that is less significant for the simulation.

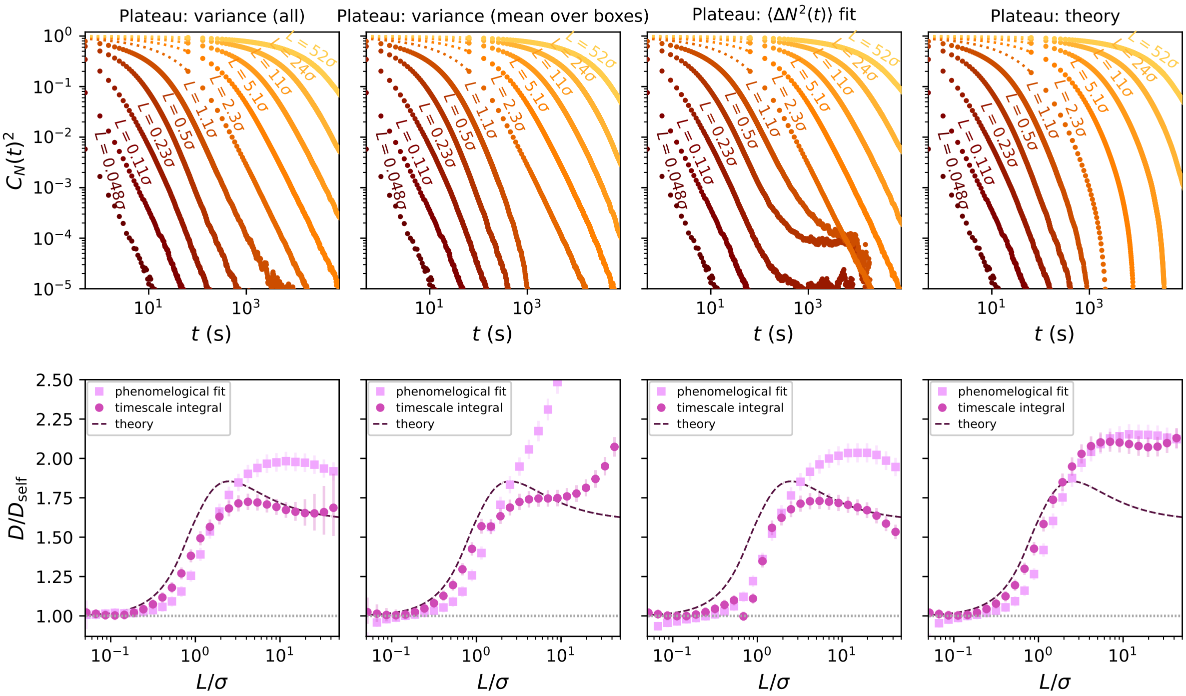

Figure S5: Curves of $C_N(t)^2$ and $D(L)$ shown for different sources of the plateau value, for simulated data at $\phi = 0.11$. "Variance (all)" is the variance of $N(t)$, calculated over all boxes and times. "Variance (mean over boxes)" is the variance over time for each box, averaged over all boxes. "$\langle \Delta N^2(t) \rangle$ fit" is a fit to Eq. S15 with $\text{Var}(N)$ and $D(L)$ free parameters. Plateau: "theory" is given by $= 2\text{Var}(N) = 2C_N(t=0)$ of Eq. (S4) of the main text. The curves for C_N come from shorter simulation runs for $L < 1.3\sigma$, and longer simulation runs for $L > 1.3\sigma$, see section 1.2.

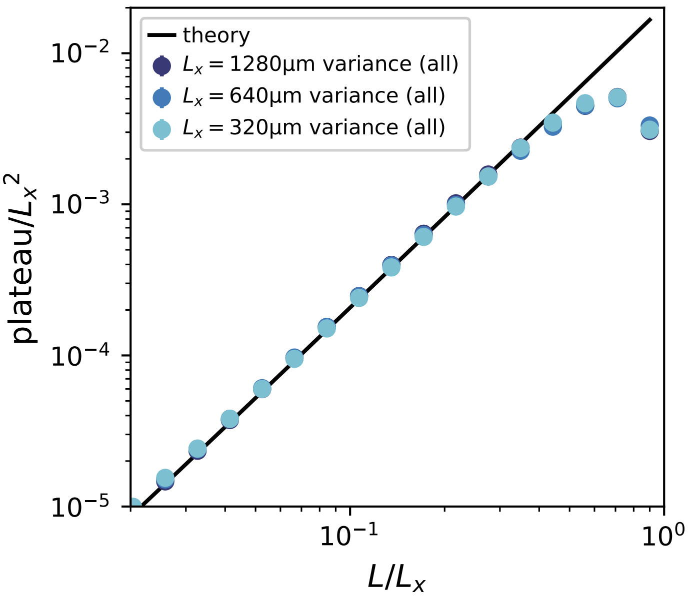

Figure S6: Effect of periodic boundary conditions in simulations on Countoscope plateau determination. The plot is similar to Fig. S4: the plateau is defined as $2\mathrm{Var}(N)$. It is either measured by computing $\mathrm{Var}(N)$ over all boxes of size L at all times (circles) or calculated via the expression of $C_N(t = 0) = \mathrm{Var}(N)$ of Eq. (S4) of the main text (solid black line). Packing fraction is $\phi = 0.11$. The different colours indicate different sizes $L_x = L_y$ of the periodic simulation domain size. The circles overlap one another for all box sizes L. The variance underestimates the theoretical plateau value for boxes $L \gtrsim 0.3L_x$, regardless of the magnitude of the simulation domain size. This is an effect of periodic boundary conditions that inhibit fluctuations at large spatial scales.

4.3 Effect of simulation box size on $D(L)$, $D(k)$

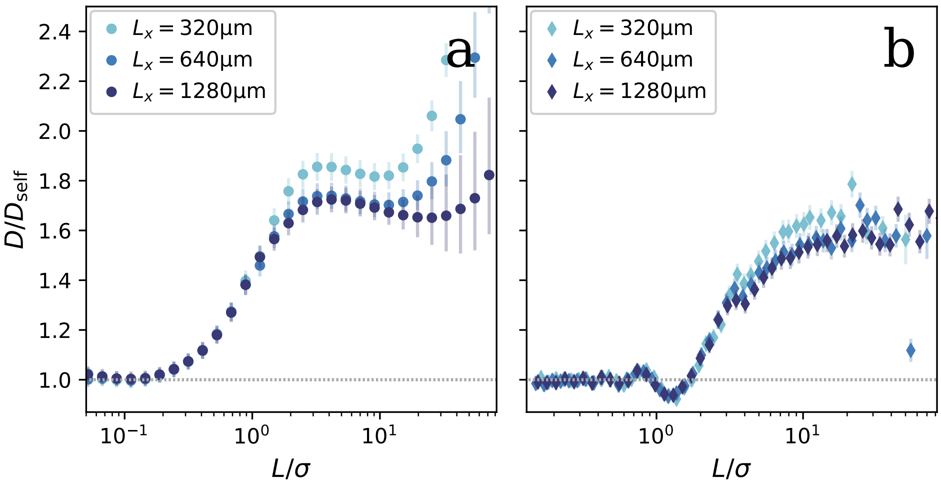

Figure S7: $D(L)$ from (a) the timescale integral (decorrelation timescale $T(L)$ of number fluctuations) and (b) dynamic structure factor for different sizes L_x of the periodic simulation domain. Increasing the domain size clearly shows that large scale $D(L)$ is reduced in larger simulation domains. In addition, the divergence seen in (a) shifts from $L \gtrsim 10\sigma$ for $L_x = 320\,\mu m$ to $L \gtrsim 30\sigma$ for $L_x = 1280\,\mu m$.

4.4 $f(k,t)$ behaviour at short and long times

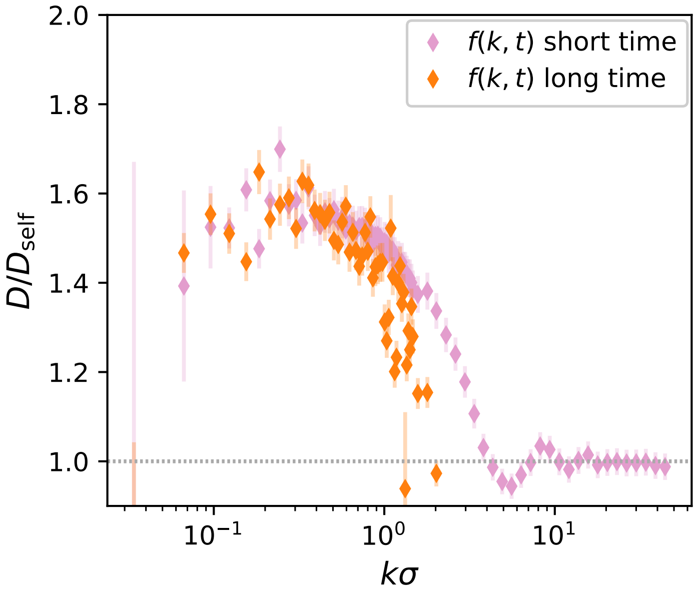

Figure S8: $D(k)$ from $f(k,t)$ at short and long times, from simulation, at $\phi = 0.11$. The dynamic structure factor decays too quickly for us to be able to extract long-time information at large k values. Short-time data is obtained from inverting the first (non-zero) point of $f(k,t)$, as described in the main text. Long-time data is obtained from fitting to Eq. 9 for $100\,\mathrm{s} < t < 1000\,\mathrm{s}$.

4.5 $f(k,t)$ short time divergence

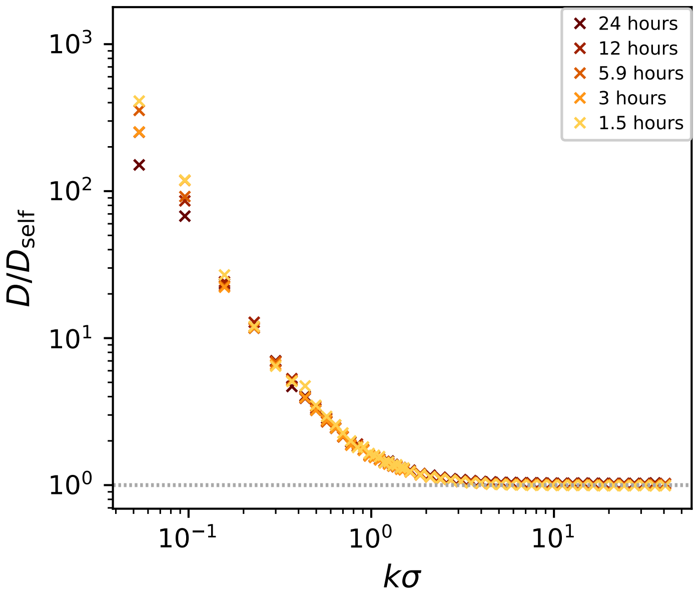

Figure S9: Experimental data of $D(k)$ at $\phi = 0.02$. The divergent behaviour of $D(k)$ at small k is independent of movie length.

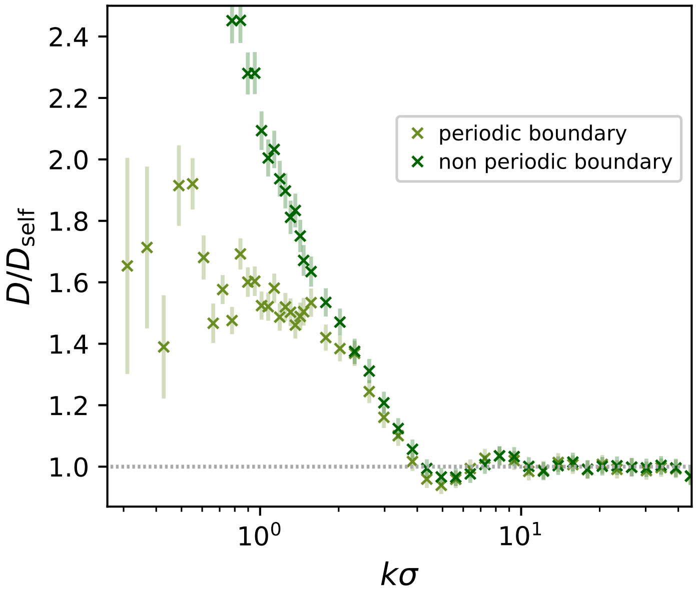

Figure S10: $D(k)$ from $f(k,t)$ from data with periodic and non periodic boundary conditions at high density (simulation, $\phi = 0.11$). Similarly to the low-density case (main text Fig. 5 d), a divergence is seen for small k (large length scales) when the data is not periodic.

Supplementary References

[1] D. S. Dean. Langevin equation for the density of a system of interacting langevin processes. *Journal of Physics A: Mathematical and General*, 29(24):L613, 1996.

[2] J. K. Dhont. *An introduction to dynamics of colloids*. Elsevier, 1996.

[3] R. Gomer. Diffusion of adsorbates on metal surfaces. *Reports on progress in Physics*, 53(7):917, 1990.

[4] J.-P. Hansen and I. R. McDonald. *Theory of simple liquids: with applications to soft matter*. Academic press, 2013.

[5] K. Kawasaki. Stochastic model of slow dynamics in supercooled liquids and dense colloidal suspensions. *Physica A: Statistical Mechanics and its Applications*, 208(1):35–64, 1994.

[6] E. K. R. Mackay, S. Marbach, B. Sprinkle, and A. L. Thorneywork. The countoscope: Measuring self and collective dynamics without trajectories. *Phys. Rev. X*, 14:041016, Oct 2024.

[7] T. H. N. Minh, B. Rotenberg, and S. Marbach. Ionic fluctuations in finite volumes: fractional noise and hyperuniformity. *Faraday Discussions*, 2023.

[8] A. L. Thorneywork, J. L. Abbott, D. G. A. L. Aarts, P. Keim, and R. P. A. Dullens. Bond-orientational order and frank's constant in two-dimensional colloidal hard spheres. *Journal of Physics: Condensed Matter*, 30(10):104003, feb 2018.

[9] A. L. Thorneywork, R. Roth, D. G. A. L. Aarts, and R. P. A. Dullens. Communication: Radial distribution functions in a two-dimensional binary colloidal hard sphere system. *The Journal of Chemical Physics*, 140(16):161106, apr 2014.

[10] A. L. Thorneywork, S. K. Schnyder, D. G. A. L. Aarts, J. Horbach, R. Roth, and R. P. A. Dullens. Structure factors in a two-dimensional binary colloidal hard sphere system. *Molecular Physics*, 116(21-22):3245–3257, nov 2018.