

# Anomalous Nanoparticle Surface Diffusion in Liquid Cell TEM is Revealed by Deep Learning-Assisted Analysis

Vida Jamali<sup>a</sup>, Cory Hargus<sup>b</sup>, Assaf Ben-Moshe<sup>a,c</sup>, Amirali Aghazadeh<sup>d</sup>, Hyun Dong Ha<sup>a</sup>, Kranthi K. Mandadapu<sup>b,e</sup>, and A. Paul Alivisatos<sup>a,c,f,g,1</sup>

<sup>a</sup>Department of Chemistry, University of California, Berkeley, CA, 94720, USA; <sup>b</sup>Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, 94720, USA; <sup>c</sup>Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA; <sup>d</sup>Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720, USA; <sup>e</sup>Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA; <sup>f</sup>Department of Materials Science and Engineering, University of California, Berkeley, CA, 94720, USA; <sup>g</sup>Kavli Energy NanoScience Institute, Berkeley, CA, 94720, USA

**The motion of nanoparticles near surfaces is of fundamental importance in physics, biology, and chemistry. Liquid cell transmission electron microscopy (LCTEM) is a promising technique for studying motion of nanoparticles with high spatial resolution. Yet, the lack of understanding of how the electron beam of the microscope affects the particle motion has held back advancement in using LCTEM for in situ single nanoparticle and macromolecule tracking at interfaces. Here, we experimentally studied the motion of a model system of gold nanoparticles dispersed in water and moving adjacent to the silicon nitride membrane of a commercial liquid cell in a broad range of electron beam dose rates. We find that the nanoparticles exhibit anomalous diffusive behavior modulated by the electron beam dose rate. We characterized the anomalous diffusion of nanoparticles in LCTEM using a convolutional deep neural network model and canonical statistical tests. The results demonstrate that the nanoparticle motion is governed by fractional Brownian motion at low dose rates, resembling diffusion in a viscoelastic medium, and continuous time random walk at high dose rates, resembling diffusion on an energy landscape with pinning sites. Both behaviors can be explained by the presence of silanol molecular species on the surface of the silicon nitride membrane and the ionic species in solution formed by radiolysis of water in presence of the electron beam.**

Liquid cell electron microscopy | Single particle tracking | Anomalous diffusion | Deep neural network

Understanding the motion of nanoparticles in boundary layers is of fundamental importance in scientific fields such as biophysics and colloidal self-assembly, and of practical importance in technological applications such as drug delivery and additive manufacturing. The physics behind the motion of nanoparticles is particularly challenging to understand due to the multitude of effects including particle-particle interactions, particle-surface interactions, and changes in the rheological properties in boundary layers close to a liquid-solid interface.

The common technique to study the motion of particles has been optical microscopy, which has limitations in terms of spatial resolution. The advent of *in situ* liquid cell transmission electron microscopy (LCTEM) has now made it possible to visualize the motion of nanoparticles near a surface with an unprecedented spatial resolution at the nanometer length scale (1–3). However, the electron beam of a transmission electron microscope (TEM), which is the key acquisition tool to enable nanoscale visualization, can significantly influence both interactions and dynamics of nanoparticles (4–6). Previous literature has reported that the motion of nanoparticles near the surface of a liquid cell and in the presence of the electron beam

is subdiffusive (*i.e.*, non-Brownian, or “anomalous”) (7–15). Such subdiffusive motion suggests that the nanoparticle motion is significantly influenced by interactions with the nearby substrate or interface, but what precisely is the nature of these interactions and the forces that create them? Are they stable or fluctuating? Do they arise because of the electron beam or are they native to the system? How do the changes in rheology within a few nanometers of the interface figure into the picture? The nature of the observed anomalies are still very much under debate as the new technique of LCTEM continues to be developed (7–15).

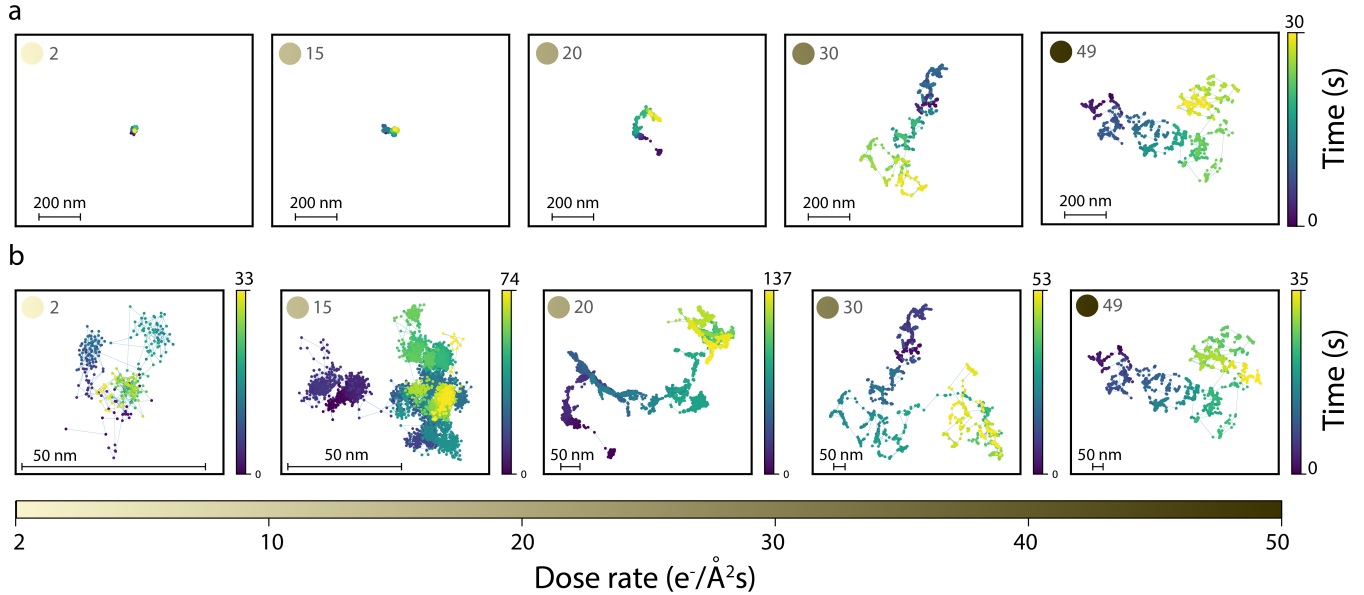
Two canonical processes that describe anomalous motion are continuous time random walk (CTRW) and fractional Brownian motion (FBM) (16–19). In the context of particle diffusion, each of these types of subdiffusive motions imply a distinct physical picture of the environment. CTRW indicates a random energy landscape of potential wells, where the time a particle spends in any well diverges when averaged over all well depths. FBM, on the other hand, indicates a viscoelastic environment such as those found in crowded fluids (20–23). The goal of this work is to identify the type of anomalous motion of nanoparticles near the surface in LCTEM, elucidate the nanoscopic physical features in the system that give rise to this motion, and understand how the electron beam can influence them.

A key challenge in studying the motion of nanoparticles under the effect of the electron beam is that one needs to resort to a limited number of short trajectories from a single *in situ* LCTEM experiment. This is because achieving high spatial resolution requires a relatively small field of view, which limits the number of nanoparticles accessible (experiments are done in dilute solutions to avoid interactions between nanoparticles). Moreover, state-of-the-art cameras on TEM microscopes are limited by lower bounds on time resolution (hundreds of frames per second) and upper bounds on measurement time (minutes long trajectories) (24). This limitation creates a challenge for canonical methods used to characterize diffusive particle dynamics such as the mean-squared displacement (MSD) analysis. These methods often rely on features of the trajectory that converge upon averaging over very long

Authors Contribution: V.J., C.H., A.B.M., A.A., H.D.H., K.K.M., and A.P.A. designed research; V.J., C.H., A.B.M., A.A., and H.D.H. performed the research; V.J., C.H., A.B.M., A.A., K.K.M., and A.P.A. analyzed data; V.J., C.H., A.B.M., A.A., H.D.H., K.K.M., and A.P.A. wrote the paper.

The authors declare no competing interests.

<sup>1</sup>To whom correspondence should be addressed. E-mail: paul.alivisatos@berkeley.edu



**Fig. 1.** Representative trajectories of 5 gold nanorods as a function of dose rate varying between 2 to 49  $e^-/\text{\AA}^2\text{s}$ , a) in their first 30 seconds and scaled to the same size showing that the diffusivity increases upon increasing the dose rate, b) diffusing for a longer time and magnified to show the change in the diffusive behavior upon increasing the dose rate.

single-particle trajectories (for systems obeying ergodicity) or hundreds of medium-length trajectories collected under the same experimental conditions (19, 25, 26). Here, we show that computational and theoretical tools can be developed to extract hidden features that exist in short trajectories of single nanoparticles in order to elucidate the type of anomalous diffusion.

In this study, we collected 30 trajectories of a model system of gold nanoparticles dispersed in water and diffusing near a silicon nitride ( $\text{SiN}_x$ ) membrane of a commercial liquid cell irradiated by a broad range of electron beam dose rates. Inspired by the recent advances in using machine learning tools to study the diffusion of single microparticles in biological media (25–28), we developed a convolutional deep neural network (CNN) model, dubbed MotionNet (MoNet), which solves an inverse problem of determining the underlying diffusion mechanism behind the anomalous motion of nanoparticles in LCTEM. The architecture of the neural network employed in MoNet is designed based on classical tests in statistics (29) and is trained on thousands of simulated short trajectories from three classes of diffusion, *i.e.* Brownian, FBM, and CTRW. Guided by MoNet, our analysis reveals that at low dose rates the anomalous diffusive motion of nanoparticles in LCTEM is governed by viscoelasticity-dominated FBM, while at high dose rates the motion is governed by a pinning site-mediated CTRW process (23). The prediction results were benchmarked against the statistical  $p$ -variation test (29) to confirm the behavior in low and high dose rate limits.

The dose rate-dependent transition can be explained by the existence of silanol molecular groups on the surface of the  $\text{SiN}_x$  membrane, which act as pinning sites and exhibit a broad distribution of restoring forces (14, 30). At low dose rates, the binding strength of these pinning sites is high compared to the thermal energy and their effective restoring force acts similar to the effect of a viscoelastic environment. This results in nanoparticle motion confined to the local vicinity of a pinning

site. Upon increasing the dose rate and thus passivating the charges on the pinning sites, the binding strength decreases, making nanoparticles more mobile, which allows them to diffuse across the  $\text{SiN}_x$  membrane only making intermittent stops on randomly distributed pinning sites. This understanding provides us with important insight into the mechanism of nanoparticle motion near a substrate in LCTEM and opens up the path to use *in situ* LCTEM as a technique for studying motions of nanoparticles in complex systems at the nanoscale.

## Results and discussions

**Anomalous diffusion of gold nanorods.** To study the effect of electron beam dose rate on the motion of nanoparticles near a surface, we chose a simple and tunable model system of 60 nm long gold nanorods (AuNRs) dispersed in water and probed their dynamics near the  $\text{SiN}_x$  membrane of a commercial TEM’s liquid cell (see Methods for details of synthesis). We collected trajectories of 30 AuNRs for electron beam dose rates ranging from 2 to 49  $e^-/\text{\AA}^2\text{s}$  (see Figure S1 for all trajectories collected). Figure 1a shows the first 30 seconds of five representative trajectories. Plotted at the same scale, these trajectories indicate that the effective diffusivity of AuNRs increases with increasing dose rate of the electron beam. Figure 1b shows the same trajectories as Figure 1a in their entirety, magnified to reveal details of the dynamics. In addition to the increase in effective diffusivity, a qualitative change in the dynamics is observed as the dose rate is increased from 2 to 49  $e^-/\text{\AA}^2\text{s}$ . At low dose rates, AuNR dynamics are dominated by motion confined to the vicinity of a local point. This motion is punctuated by infrequent, relatively long-distance jumps. At high dose rates, long-distance jumps between short periods of confinement become the dominant behavior at observation timescales.

To identify the underlying diffusive behavior and to understand how the electron beam changes the local environment

and the local interactions, we first calculated the MSD. In the framework of anomalous diffusion, the MSD is described by the power law (19):

$$\langle \delta x(t)^2 \rangle \sim D_\alpha t^\alpha. \quad [1]$$

Here, brackets denote an ensemble average, and accordingly we refer to this as the ensemble averaged or e-MSD. If  $\alpha = 1$  the process is characterized by Brownian motion, and if  $\alpha < 1$  or  $\alpha > 1$  the process is subdiffusive or superdiffusive, respectively. The MSD may also be computed by window averaging over a single trajectory, which we refer to as a time averaged or t-MSD, and is defined by

$$\overline{\delta x^2(\Delta)} = \frac{1}{T - \Delta} \int_0^{T-\Delta} dt (x(t) - x(t + \Delta))^2. \quad [2]$$

Here,  $T$  is the total measurement time,  $\Delta$  is the time delay window, and  $\langle \cdot \rangle$  indicates an average over time. For ergodic processes, the e-MSD and t-MSD are equal in the long time limit as  $T \rightarrow \infty$ . In case of non-ergodic subdiffusive processes, the e-MSD contains more information about the underlying anomaly mechanism; however, it is not practically accessible in many experimental systems, including LCTEM as it is available today. The t-MSD measurements as a function of  $\Delta$  for all thirty trajectories are presented in Figure S2. We also measured the related time-averaged diffusion constant  $D_\alpha = \overline{\delta x^2(\Delta)}/\Delta^\alpha$ , using Eq. [2], which varies between 10 and  $10^4$  nm<sup>2</sup>/s for values of  $\Delta \leq 0.25$  s and for all dose rates studied across 30 trajectories in three experiments; see Figure S3. This shows that the motion of AuNRs near the surface is orders of magnitude slower than what is theoretically estimated for a Brownian nanoparticle in bulk water outside of TEM based on the Stokes-Einstein relation ( $D = k_B T / (6\pi\eta L) \approx 4 \times 10^6$  nm<sup>2</sup>/s with  $\eta$  the viscosity of the medium and  $L$  the characteristic size of the diffusing nanoparticle). The slow motion of AuNRs observed here is consistent with previous reports on the suppressed diffusive motion of nanoparticles in LCTEM experiments (8, 9, 11, 13, 14). Figures S2 and S3 also show that  $D_\alpha$  increases as the dose rate is increased, consistent with the observations from Figure 1a. However, it is not possible to identify the type of diffusion as well as whether it is anomalous or not based solely on the t-MSD curves. This can be explained by a closer look at two common anomalous diffusion models, subdiffusive CTRW and subdiffusive FBM, and their corresponding MSDs.

**Models of anomalous diffusion.** Diffusion processes in which particles move with stop-and-go motion on an energy landscape with heterogeneous pinning sites are well described as a continuous time random walk (CTRW) (19). In a CTRW process, a particle moves by making random jumps in space and time (see SI Appendix for details). The particle remains immobile for a random “waiting time”  $\tau$ , drawn from distribution  $\psi(\tau)$ , before jumping in the distance and direction  $\Delta \mathbf{x}$ , drawn from the distribution  $\lambda(\Delta \mathbf{x})$  (22). If  $\psi(\tau)$  is heavy-tailed, *i.e.*, the asymptotic behavior at large  $\tau$  decays as  $\psi(\tau) \sim 1/\tau^{1+\alpha}$  with  $0 < \alpha < 1$ , the mean waiting time  $\langle \tau \rangle$  diverges ( $\langle \tau \rangle \rightarrow \infty$ ) and the resulting process is subdiffusive (21). The diverging  $\langle \tau \rangle$  also indicates that ergodicity is broken; no matter how long the measurement time  $T$  is, the t-MSD and the ensemble averaged t-MSD (average of t-MSDs over an ensemble of particles, or et-MSD) will not be the same (21). It can be shown

that the et-MSD for a CTRW process can be written as (see SI Appendix for mathematical derivation) (31, 32)

$$\langle \overline{\delta x^2(\Delta)} \rangle \sim D_\alpha \frac{\Delta}{T^{1-\alpha}}. \quad [3]$$

Eq. [3] shows that for a CTRW process, the et-MSD is a linear function of time delay,  $\Delta$ . The e-MSD is obtained from the et-MSD in the limit  $\Delta \rightarrow T$ , recovering the anomalous form of Eq. [1]. This property of the subdiffusive CTRW process makes it extremely difficult to identify and to estimate its inherent  $\alpha$  value, when only a limited number of short trajectories from an experiment is accessible since no anomaly can be detected by measuring the t-MSD.

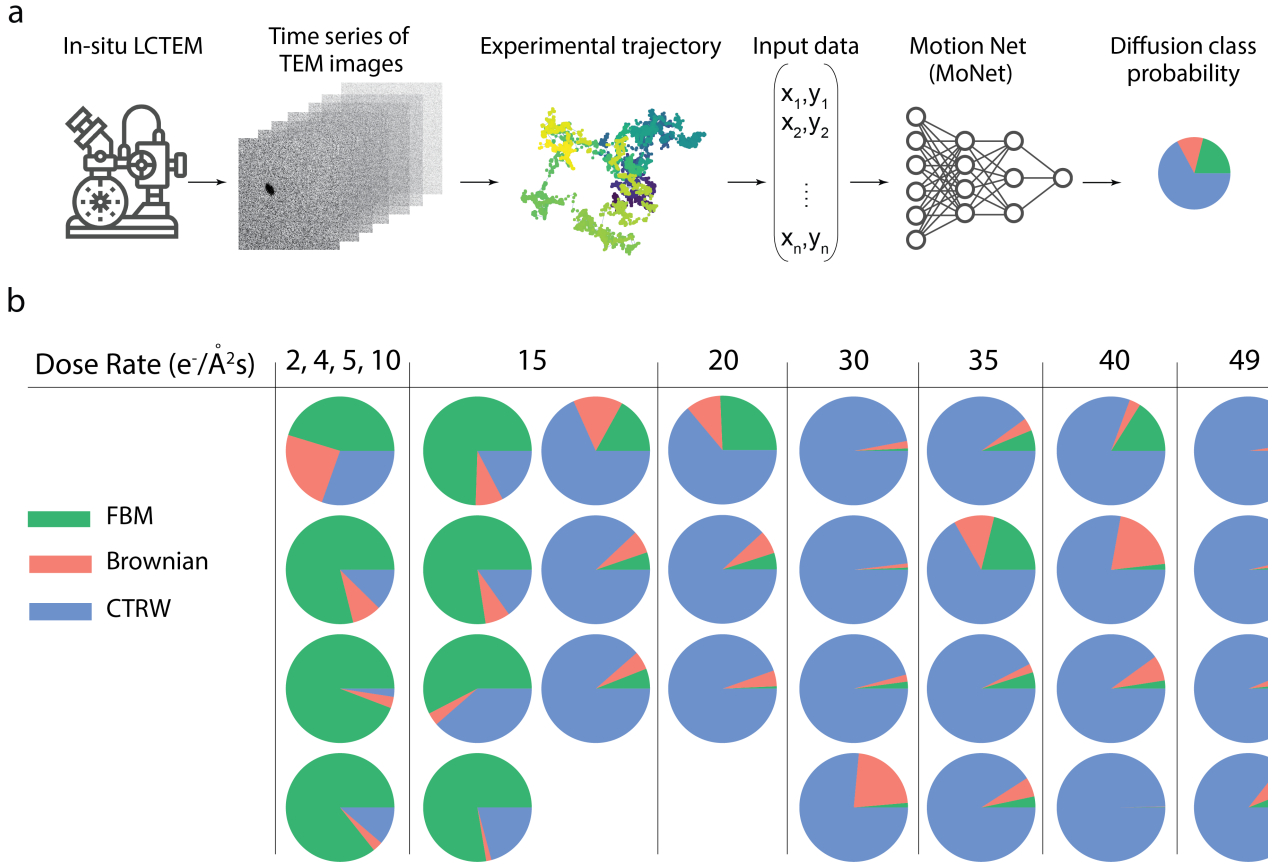
Another canonical model of subdiffusion is Fractional Brownian Motion (FBM) (17). Subdiffusive FBM can be qualitatively described as a random process in which the direction of each step is anti-correlated with the previous step, resulting in the next step having a higher probability than random to be in the opposite direction (30). This correlation of positions at two different points in time,  $t_1$  and  $t_2$  along the trajectory can be expressed as:

$$\langle x(t_1)x(t_2) \rangle = D_\alpha (|t_1|^\alpha + |t_2|^\alpha - |t_1 - t_2|^\alpha), \quad [4]$$

where as before,  $\alpha < 1$  corresponds to subdiffusion. Unlike the CTRW model, a FBM process is ergodic, and thus t-MSD and its ensemble average, *i.e.* et-MSD, are the same and follow  $\langle \overline{\delta x^2(\Delta)} \rangle = D_\alpha \Delta^\alpha$  (see SI Appendix for mathematical derivation of a FBM process).

**Deep learning analysis.** To identify the underlying anomalous diffusion process for a limited set of short trajectories in a LCTEM experiment, we developed a convolutional neural network model which we have named MoNet (shown in Figure 2a). We trained MoNet on 10,000 simulated trajectories from three classes of diffusion: Brownian, subdiffusive FBM, and subdiffusive CTRW. Each simulated trajectory was 300 frames in length; short enough to cover the shortest experimental trajectories collected and long enough to achieve more than 90% validation accuracy (see Figure S6). For consistency, the model was then applied to 300-frame intervals of all nanoparticle trajectories. The final results are reported as the predicted probability for each diffusion class, averaged over the entire length of the trajectory (see Figure S6 for validation accuracy on an independent set of test data). As shown in Figure 2a, MoNet receives input data in the form of a matrix comprising the  $x$  and  $y$  coordinate of the locations of the nanoparticle throughout the trajectory and outputs the probability of the predicted diffusion class, *i.e.* FBM, CTRW, and Brownian. The architecture of MoNet is inspired by previous literature for temporal sequence type data such as particle trajectories (26, 33). See the Methods section and SI Appendix for details of the architecture of MoNet.

Figure 2b presents the predicted probability of the diffusion class for all 30 trajectories as a function of dose rate, increasing from top to bottom in the first column and left to right across the table; see Figure S7 for the probability values associated with each class. Interestingly, there is a crossover from FBM to CTRW as the electron beam dose rate is increased, consistent with the qualitative picture of Figure 1b. The crossover occurs around the dose rate of  $15 \text{ e}^-/\text{\AA}^2\text{s}$ , where 7 trajectories have been collected.



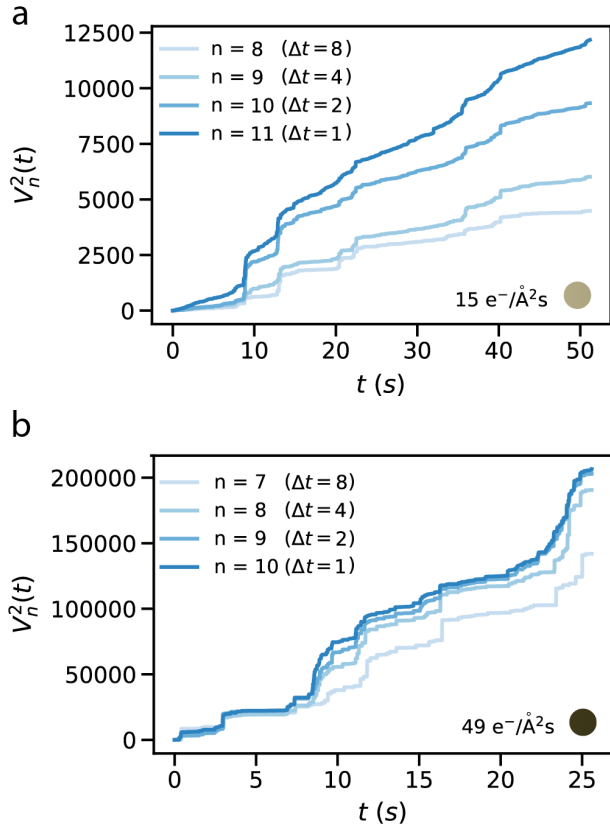
**Fig. 2.** a) Deep neural network pipeline for anomalous diffusion classification on *in situ* LCTEM data using our nanoparticle motion classifier MoNet. b) Neural network analysis results for all trajectories studied as a function of dose rate (increasing from top to bottom in the first column and left to right across the table). Pie charts show the diffusion class probability where at low dose rates, there is a higher probability associated with a fractional Brownian motion (FBM) (green) and at high dose rates there is a higher probability associated with a continuous time random walk (CTRW) (blue).

To verify the results we compared MoNet against a statistical method, known as the  $p$ -variation test,  $V_n^p$  (see the Methods section for definition).  $p$ -variation has been successful in distinguishing FBM from CTRW for medium length trajectories (29). Here, we have analyzed the quadratic variation ( $p$ -variation for  $p = 2$ ), which measures the sum of squares of increments of a trajectory of length  $T = 2^{N_{max}}$ , divided into  $2^n$  segments. For a FBM process, the quadratic variation must diverge as  $n \rightarrow \infty$  (*i.e.*, the size of time increment  $\Delta t \rightarrow 1$  frame), while for a CTRW process, the quadratic variation must stabilize with increasing  $n$  (29, 34). Comparison of our predictions with the quadratic variation results presented in Figure S8 confirms that there is indeed a crossover from FBM to CTRW while increasing the dose rate. Figure 3 shows the quadratic variation results for two example trajectories of Figure 1b at dose rates 15 and 49  $e^-/\text{\AA}^2\text{s}$ . The unbounded increase in the slope of the quadratic variation vs. measurement time curve as  $n \rightarrow N_{max}$  (*i.e.*,  $\Delta t \rightarrow 1$  frame) confirms that at dose rate 15  $e^-/\text{\AA}^2\text{s}$ , the trajectory is predominantly characterized by a FBM behavior. However, for a higher dose rate of 49  $e^-/\text{\AA}^2\text{s}$ , the quadratic variation curve does not show any specific dependence as  $n \rightarrow N_{max}$ , suggesting that the anomaly does not stem from a FBM process.

Another characteristic of FBM and CTRW processes in terms of displacement,  $\delta x$ , is their probability distribution of displacements  $P(\delta x)$  (19, 30). Comparison of the distribution

of displacements collected over time delays of 0.0125 s in Figure 4a for two example trajectories at dose rates 15 and 49  $e^-/\text{\AA}^2\text{s}$ , (same trajectories as Figure 3), also confirms the presence of a FBM process at low dose rates with a Gaussian distribution and a CTRW process at high dose rates with a power-law tailed distribution. The power-law exponent of this tail is estimated to be about  $-2.0$  (see Figure 4a). Figure S9 shows that this power-law value of  $-2.0$  is consistent for all high-dose rate trajectories studied here.

The power-law decay of the probability distribution  $P(\delta x)$  for large values of displacement,  $\delta x$ , at high dose rates does not necessarily mean that the underlying CTRW process is subdiffusive (35). However, it suggests that there is a broad distribution of binding sites on the surface of the  $\text{SiN}_x$  membrane. It is known that for harmonic energy potentials with equal binding stiffness  $k$ , the resulting probability distribution of displacements must follow a Gaussian form,  $P(\delta x) = \exp(-k(\delta x)^2/k_B T)$  (30, 36). Hence, the non-Gaussian and heavy-tailed probability distribution of displacements observed for all high-dose rate trajectories indicates that binding sites with various binding affinities exist over the surface of the  $\text{SiN}_x$  membrane, suggesting that an underlying CTRW process could be subdiffusive. To confirm that the CTRW process observed at high dose rates is subdiffusive, we used MoNet trained on three thousand simulated CTRW trajectories with  $\alpha$  values between 0.1 and 0.99, and predicted



**Fig. 3.** Quadratic-variation test results vs. measurement time,  $t$ , for trajectories collected at dose rates a) 15 and b)  $49 \text{ e}^-/\text{\AA}^2\text{s}$  selected from Figure 1b. a) The slope of the quadratic-variation curve increases as  $n$  increases (*i.e.*, size of time increments  $\Delta t$  decreases to 1 frame) indicative of a FBM process, b) the slope of the quadratic-variation curve converges as  $n$  increases, indicative of a CTRW behavior at high dose rates.

the  $\alpha$  exponent for all trajectories collected (see Figure S10). The results show that the underlying mechanism at high dose rates is subdiffusive with  $\alpha$  exponents ranging from 0.7 – 0.8.

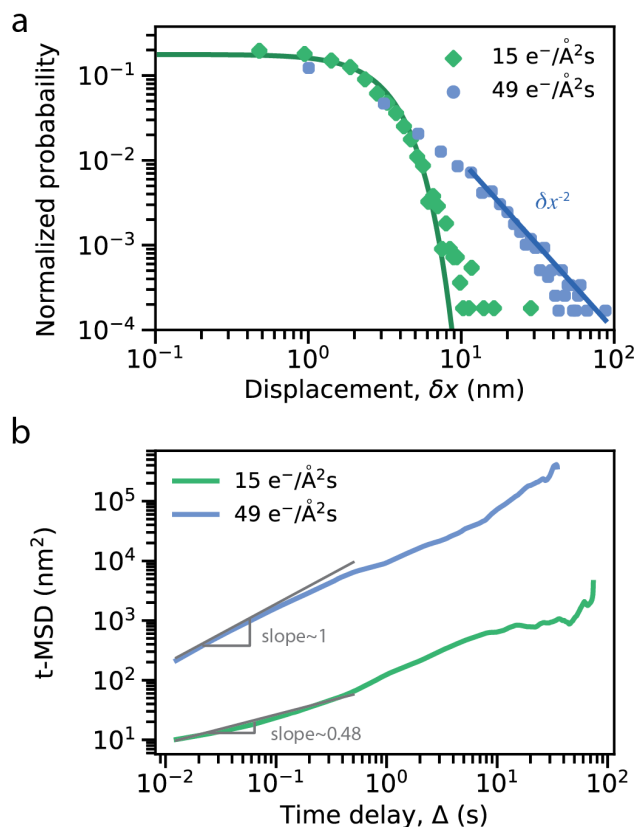
We also did a similar analysis using MoNet to predict  $\alpha$  exponents of FBM processes (commonly known as the Hurst exponent  $H = \alpha/2$  in the literature (37)); see SI Appendix and Figure S10. The  $\alpha$  exponent obtained from this analysis is very similar to the values of  $\alpha$  extracted from t-MSD curves (see Figure S10). Figure 4b shows the t-MSD curves calculated for trajectories of dose rates 15 and  $49 \text{ e}^-/\text{\AA}^2\text{s}$  with  $\alpha$  exponents of 0.48 and 1, respectively. As shown in Eq. [3], the t-MSD curve of a CTRW process grows linearly in time delay  $\Delta$ , consistent with our measurements shown in Figure 4b. For subdiffusive FBM processes, the t-MSD curve grows sublinearly in time delay  $\Delta$ . Therefore, the t-MSD curve can only provide us with a value of  $\alpha$  at low dose rates, where the process is predominantly characterized by a FBM model and et-MSD measurements are further required to estimate the value of  $\alpha$  for CTRW processes. Using MoNet predictions for the  $\alpha$  exponent for both low and high dose rate trajectories, we showed that at all dose rate studied the underlying diffusive process is subdiffusive.

**Nanoscope interpretation.** The physical picture governing the different diffusive behavior at low and high electron beam dose rates may be explained by the molecular groups existing on

the surface of the  $\text{SiN}_x$  membrane of the TEM liquid cell, which in turn are influenced by the electron beam. It has been previously suggested that the surface of the  $\text{SiN}_x$  membrane is decorated with silanol  $\text{Si}-\text{O}$  molecular groups (14). These silanol groups that are randomly distributed across the membrane create pinning sites and can locally trap nanoparticles, which are positively charged with cetyltrimethylammonium chloride (CTAC) ligands. At low electron beam dose rates ( $\leq 15 \text{ e}^-/\text{\AA}^2\text{s}$ ), the thermal energy of AuNRs is smaller than the binding strength of these pinning sites resulting in particles being trapped for significant periods in the vicinity of a local pinning site, with membrane restoring forces and solvent interactions acting as a viscoelastic medium. This viscoelastic picture may be explained by the hydrogen bonding of the water molecules with the  $\text{Si}-\text{O}$  species on the surface of the  $\text{SiN}_x$  that may result in a gel-like viscoelastic water layer next to the membrane at low dose rates, leading to the FBM behavior at this dose rate. We note that a close inspection of the trajectories of Figure 1b and Figure S1 shows that for some of our low dose rate experiments, the immobility in pinning sites is punctuated by a few relatively long-distanced jumps. Yet, these jumps are smaller than 50 nm which is smaller than the body length of the AuNRs studied ( $\sim 60 \text{ nm}$ ) and could be explained as the head or tail of the same AuNR being trapped in the same pinning site. Upon increasing the dose rate, radiolysis of water occurs, changing the local pH value of the solution close to the membrane. This change in pH results in free  $\text{H}^+$  ions in solution that passivate the  $\text{Si}-\text{O}$  groups on the membrane's surface, reducing their binding strength. Therefore, at high dose rates, AuNRs can occasionally de-trap and move with long distance jumps until they get trapped in another pinning site associated with a waiting time  $\tau$  drawn from a heavy tailed distribution function  $\psi(\tau)$ .

While this interpretation could explain this set of observations, we note that alternative scenarios may exist such as coexistence of both FBM and CTRW behavior. A close look at the predictions of Figure 2b shows that for certain low dose rates (see  $15 \text{ e}^-/\text{\AA}^2\text{s}$ ) classified as FBM by MoNet, there is a non-negligible probability associated with the CTRW class. By tracking the predicted diffusion class along the entire length of the trajectory for each window of 300 frames separately, we can observe that windows including long distance jumps of  $> 50 \text{ nm}$  are more likely to be classified as CTRW (see Figure S12).

The presence of these jumps, even at low dose rates, also shows up in the t-MSD curves. The t-MSD curve for the dose rate of  $15 \text{ e}^-/\text{\AA}^2\text{s}$  presented in Fig. 4b has an  $\alpha$  exponent of 0.48 at short time delays, while at long time delays the exponent increases to 1. This is in contrast to high dose rates, where throughout the t-MSD curve, the  $\alpha$  exponent remains constant at a value of 1. The change in the  $\alpha$  exponent as well as the non-negligible probability associated with the CTRW class at dose rates of  $15 \text{ e}^-/\text{\AA}^2\text{s}$  suggests that both FBM and CTRW behavior could potentially coexist at this dose rate but at different timescales. Furthermore, the scatter in t-MSD curves for all dose rates in Figure S2 shows that that ergodicity might be broken even at low dose rates where the diffusion class is predominantly characterized by FBM, which is by definition an ergodic process. This is reminiscent of the subordinated diffusion processes reported in biological systems as well as single molecule tracking experiments in water (30, 38–40). This



**Fig. 4.** a) Probability distribution of the absolute value of displacement for trajectories collected at the dose rates 15 and 49  $e^-/\text{\AA}^2\text{s}$  for time intervals of 0.0125 s, compared to a Gaussian fit (solid green line) and a power-law tail fit with power-law exponent of  $-2$  (solid blue line). b) Time averaged mean squared displacement vs. time delay calculated for the same trajectories of panel (a). Solid gray lines show the fit to the MSD curves at short time delays with the slope of 0.48 for the low dose rate and slope of 1 for the high dose rate.

type of subordinated diffusion is complex to capture through canonical methods and indeed requires data spanning multiple timescales both on short and long time delays. Therefore, while the current data is insufficient to support or nullify this hypothesis (especially at high dose rates), our analysis suggests the possibility of such a scenario. Regardless, the presence of predominantly FBM behavior at low dose rates and CTRW at high dose rates supports the interpretation that the diffusive motion at low dose rates is mostly influenced by the local viscoelasticity of the fluid next to the surface and at high dose rates the motion is governed by the heterogeneous pinning sites in the timescales studied ( $\Delta = 0.01$  s to 100 s). Therefore, the electron beam dose rate not only increases the diffusion coefficient, but also it fundamentally alters the fluid and the dominant diffusive behavior of nanoparticles near the membrane.

This understanding of how the electron beam can affect the local environment near the membrane, which in turn governs the diffusive motion of nanoparticles near the surface, can be used in applications of nanoparticles in LCTEM as nanoscale probes to study the local material properties of the fluid near the surface. Similar analysis can also be performed on undamped motion of nanoparticles in bulk in LCTEM, (41, 42) to investigate the effect of electron beam on the bulk material

properties. Furthermore, the change in the local material properties of the fluid next to the surface in presence of the electron beam may play a role in other LCTEM studies such as *in situ* growth of nanocrystals (43–45). The knowledge base developed here can be also extended to study the motion of nanoparticles in LCTEM near surfaces with various combinations of nanoparticles, fluids, and surfaces with high spatial resolution.

## Materials and Methods

**Chemicals and materials.** Hexadecyltrimethylammonium bromide (CTAB, > 98.0%), hexadecyltrimethylammonium chloride (CTAC, > 95.0%) and sodium oleate (NaOL, > 97.0%) were purchased from TCI America. Acetone (99.5%) was purchased from Fisher Scientific. Hydrogen tetrachloroaurate trihydrate ( $\text{HAuCl}_4 \cdot 3\text{H}_2\text{O}$ ,  $\geq 99.9\%$ ), L-ascorbic acid (BioXtra,  $\geq 99.0\%$ ), silver nitrate ( $\text{AgNO}_3$ ,  $\geq 99.0\%$ ), sodium borohydride ( $\text{NaBH}_4$ , 99.99%), and hydrochloric acid (36.5%–38.0% wt.%) were obtained from Sigma Aldrich (USA).  $\text{NaBH}_4$  powder was stored in an argon glovebox.  $\text{HAuCl}_4 \cdot 3\text{H}_2\text{O}$ , L-ascorbic acid, and  $\text{AgNO}_3$  were stored in a vacuum desiccator at room temperature. Deionized water (DI-water, Milipore, Milford, MA, USA) was used for all aqueous solutions. All the glassware was thoroughly cleaned using freshly prepared aqua regia (3 : 1 volume ratio of HCl and  $\text{HNO}_3$ , respectively) followed by fully rinsing with copious amounts of DI-water. All chemicals were of reagent grade and used without further purification.

**Liquid cell preparation.** Commercially available silicon nitride liquid cell top (EPT-52W-10) and bottom (EPB-52DNS) microchips (Protochips Inc.) with electron transparent membranes and a 150 nm static spacer, were cleaned by being immersed in a clean petri dish filled with acetone to remove the protective resist coating and immediately transferred to a second petri dish filled with high purity ethanol. The microchips were then dried by blotting them on a filter paper to remove the excess ethanol. The microchips were fully dried by blowing gently nitrogen gas parallel to their surface. Following that they were plasma-treated for 3 minutes to remove any residual organic material and to improve their surface hydrophilicity. The microchips were then assembled in a Poseidon 200 holder according to the Protochips Inc. protocols with 0.75  $\mu\text{L}$  of the AuNR solution containing an extra 5 mM of CTAC ligands.

**TEM imaging.** *in situ* experiments were performed on a FEI Tecnai T20 S-TWIN TEM operating at 200 KV with a LaB6 filament. Time series of images were collected using a Gatan Rio 16 IS camera in Digital Micrograph format at nominal magnifications of 25.3 kx and 38.1 kx with various exposures of 0.1, 0.05, 0.0125, and 0.00625 seconds corresponding to frame rates of 10, 20, 80, and 160 frames per seconds with  $4096 \times 4096$ ,  $2048 \times 2048$ , and  $1024 \times 1024$  pixels by pixels readout, resulting in 0.355208, 0.710415, and 1.42083 nm/pixels resolutions, respectively. Prior to imaging, the electron beam dose rate was calibrated at each magnification using a custom digital micrograph script as described in the previous literature by converting counts to electrons with a conversion value of 124 (46). The range of dose rates accessible after calibration at this magnification spans from 2 to 49  $e^-/\text{\AA}^2\text{s}$ . Data were collected in three sets of experiments using the same dose rates to assure the consistency of the outcomes. Furthermore, dose rates were increased and decreased to verify the reversibility of the process. Time series of high dimensional images were processed in MATLAB using custom scripts to obtain trajectories of nanoparticles presented in Figure S1 by tracking the centroid of AuNRs in each frame.

**Synthesis of gold nanorods.** Homogenous AuNRs were synthesized by a facile seed-mediated growth involving a binary surfactant mixture (47). The seed solution was prepared as follows: 10 mL of 0.1 M CTAB solution was mixed with 100  $\mu\text{L}$  of 25 mM  $\text{HAuCl}_4$  in a 20 mL scintillation vial under vigorous stirring. 600  $\mu\text{L}$  of ice cooled 10 mM  $\text{NaBH}_4$  was rapidly injected into the Au-CTAB solution and stirred for 2 minutes. Upon the addition of  $\text{NaBH}_4$ , the color of the seed solution turned yellow-brownish. Afterward,

the seed solution was left undisturbed at 28°C for 30 minutes prior to use in the following step.

The growth solution was obtained by first mixing 3.6 g of CTAB and 0.4936 g of NaOL in 196 mL of DI-water in a 500 mL Erlenmeyer flask. The solution was heated with occasional agitation until all the CTAB was dissolved. The mixture was allowed to cool down to 30°C and 1.45 mL of 10 mM AgNO<sub>3</sub> was then added under stir at 700 rpm for 15 min. Afterward, 4 mL of 25 mM HAuCl<sub>4</sub> was added to the mixture and kept undisturbed at 28°C for 90 min. The yellowish color of growth solution turned to colorless. 840 μL of HCl was added to the solution and the mixture was stirred at 400 rpm for 15 min. Finally, 500 μL of 0.064 M ascorbic acid was injected into the growth solution, and the mixture was vigorously stirred at 1200 rpm for 30 s. 80 μL of the seed solution was then injected, and the solution was stirred for 30 s before left undisturbed at 28°C for 12 hr to complete the growth process. 40 mL of the final products were isolated by centrifugation at 8,000 rpm for 15 min followed by careful removal of the supernatant. 50 mL of DI-water was added to the pellet and the mixture was sonicated briefly to disperse the pellet for long-term storage. For the sample preparation of the liquid cell experiment, a second centrifugation step was performed at 5,500 rpm for 10 min followed by removal of the supernatant and adding 50 mL of DI-water. 1 mL of the stock solution was centrifuged at 5,500 rpm for 8 min and the supernatant was carefully removed. 1 mL of 50 mM CTAC solution was added and sonicated for 10 min. The solution was centrifuged again at 5,500 rpm for 8 min followed by removal of the supernatant and adding 1 mL of the DI-water.

**Deep learning.** MoNet architecture consists of 6 convolution layers (including 5 dilated convolution layers) followed by 3 dense layers. The dilated layers have 32 filters of sizes  $k = 2, 3, 4, 10$ , and 20 with a combination of dilation factors of  $2^n$  for  $n = 0, 1, 2$ , and 3 (inspired by  $p$ -variation method) to capture long distance correlations existing in increment of  $2^n$  along the trajectory. See Figures S4 and S5 for the schematic of the neural net architecture. The validation accuracy of MoNet has been tested on simulated trajectories of different length (Figure S6). For a 300-frame long trajectory the prediction accuracy of the diffusion class is 90%. The mean squared error associated with the task of  $\alpha$  prediction in CTRW and FBM models are 0.02 and 0.003, respectively. See SI Appendix for more details.

**$p$ -variation test.** To distinguish between subdiffusive FBM and CTRW dynamics, Magdziarz et al. proposed the  $p$ -variation test (29, 34). This test generalizes the concept of the total variation  $V$ , in which the increments (*i.e.*, particle displacements) are summed over the entire trajectory. The  $p$ -variation  $V_n^{(p)}(t)$  generalizes the concept of total variation by exponentiating each increment by  $p$  before summing (48)

$$V_n^{(p)}(t) = \sum_{j=1}^{(2^n)t} |x(j/2^n) - x((j-1)/2^n)|^p. \quad [5]$$

Given a trajectory with a length of  $2^N$ , in case of  $p = 2$  (quadratic variation)  $V_n^2(t)$ , we sum up the square of the increments which are spaced  $2^{N-n}$  in time. See SI Appendix and Figure S4a for more details.

## References.

- N de Jonge, FM Ross, Electron microscopy of specimens in liquid. *Nat. Nanotechnol.* **6**, 695–704 (2011).
- FM Ross, Opportunities and challenges in liquid cell electron microscopy. *Science* **350** (2015).
- FM Ross, *Liquid Cell Electron Microscopy*. (Cambridge University Press), (2016).
- TJ Woehl, et al., Experimental procedures to mitigate electron beam induced artifacts during in situ fluid imaging of nanomaterials. *Ultramicroscopy* **127**, 53–63 (2013).
- DM Marolf, MR Jones, Measurement challenges in dynamic and nonequilibrium nanoscale systems. *Anal. Chem.* **91**, 13324–13336 (2019).
- H Wu, H Friedrich, JP Patterson, NA Sommerdijk, N de Jonge, Liquid-phase electron microscopy for soft matter science and biology. *Adv. Mater.*, 2001582 (2020).
- HM Zheng, SA Claridge, AM Minor, AP Alivisatos, U Dahmen, Nanocrystal diffusion in a liquid thin film observed by in situ transmission electron microscopy. *Nano Lett.* **9**, 2460–2465 (2009).

- Q Chen, et al., 3D motion of DNA-Au nanoconjugates in graphene liquid cell electron microscopy. *Nano Lett.* **13**, 4556–4561 (2013).
- J Lu, Z Aabdin, ND Loh, D Bhattacharya, U Mirsaidov, Nanoparticle dynamics in a nanodroplet. *Nano Lett.* **14**, 2111–2115 (2014).
- TJ Woehl, T Prozorov, The mechanisms for nanoparticle surface diffusion and chain self-assembly determined from real-time nanoscale kinetics in liquid. *J. Phys. Chem. C* **119**, 21261–21269 (2015).
- Q Chen, et al., Interaction potentials of anisotropic nanocrystals from the trajectory sampling of particle motion using in situ liquid phase transmission electron microscopy. *ACS Cent. Sci.* **1**, 33–39 (2015).
- A Verch, M Pfaff, N de Jonge, Exceptionally slow movement of gold nanoparticles at a solid/liquid interface investigated by scanning transmission electron microscopy. *Langmuir* **31**, 6956–6964 (2015).
- LR Parent, et al., Tackling the challenges of dynamic experiments using liquid-cell transmission electron microscopy. *Accounts Chem. Res.* **51**, 3–11 (2018).
- SW Chee, U Anand, G Bisht, SF Tan, U Mirsaidov, Direct observations of the rotation and translation of anisotropic nanoparticles adsorbed at a liquid-solid interface. *Nano Lett.* **19**, 2871–2878 (2019).
- E Bakalis, et al., Complex nanoparticle diffusional motion in liquid cell transmission electron microscopy. *The J. Phys. Chem. C* **124**, 14881–14890 (2020).
- EW Montroll, GH Weiss, Random walks on lattices. ii. *J. Math. Phys.* **6**, 167–181 (1965).
- BB Mandelbrot, JW Van Ness, Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422–437 (1968).
- R Metzler, J Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Reports* **339**, 1–77 (2000).
- R Metzler, JH Jeon, AG Cherstvy, E Barkai, Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* **16**, 24128–24164 (2014).
- I Goychuk, Viscoelastic subdiffusion: From anomalous to normal. *Phys. Rev. E* **80**, 046125 (2009).
- E Barkai, Y Garini, R Metzler, of single molecules in living cells. *Phys. Today* **65**, 29 (2012).
- IM Sokolov, Models of anomalous diffusion in crowded environments. *Soft Matter* **8**, 9043–9052 (2012).
- Y Meroz, IM Sokolov, J Klafter, Test for determining a subdiffusive model in ergodic systems from single trajectories. *Phys. Rev. Lett.* **110**, 090601 (2013).
- N de Jonge, L Houben, RE Durin-Borkowski, FM Ross, Resolution and aberration correction in liquid cell transmission electron microscopy. *Nat. Rev. Mater.* **4**, 61–78 (2019).
- S Bo, F Schmidt, R Eichhorn, G Volpe, Measurement of anomalous diffusion using recurrent neural networks. *Phys. Rev. E* **100**, 010102 (2019).
- N Granik, et al., Single-particle diffusion characterization by deep learning. *Biophys. Journal* **117**, 185–192 (2019).
- G Muñoz-Gil, MA Garcia-March, C Manzo, JD Martín-Guerrero, M Lewenstein, Single trajectory characterization via machine learning. *New J. Phys.* **22**, 013010 (2020).
- F Cichos, K Gustavsson, B Mehlig, G Volpe, Machine learning for active matter. *Nat. Mach. Intell.* **2**, 94–103 (2020).
- M Magdziarz, A Weron, K Burnecki, J Klafter, Fractional Brownian motion versus the continuous-time random walk: A simple test for subdiffusive dynamics. *Phys. Rev. Lett.* **103**, 180602 (2009).
- R Sarfati, DK Schwartz, Temporally anticorrelated subdiffusion in water nanofilms on silica suggests near-surface viscoelasticity. *ACS Nano* **14**, 3041–3047 (2020).
- Y He, S Burov, R Metzler, E Barkai, Random time-scale invariant diffusion and transport coefficients. *Phys. Rev. Lett.* **101**, 058101 (2008).
- S Burov, JH Jeon, R Metzler, E Barkai, Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking. *Phys. Chem. Chem. Phys.* **13**, 1800–1812 (2011).
- S Bai, JZ Kolter, V Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).
- M Magdziarz, J Klafter, Detecting origins of subdiffusion:  $p$ -variation test for confined systems. *Phys. Rev. E* **82**, 011129 (2010).
- B Wang, SM Anthony, SC Bae, S Granick, Anomalous yet Brownian. *Proc. Natl. Acad. Sci.* **106**, 15160–15164 (2009).
- M Lindner, G Nir, A Vivante, IT Young, Y Garini, Dynamic analysis of a diffusing particle in a trapping potential. *Phys. Rev. E* **87**, 022716 (2013).
- A Weron, K Burnecki, S Mercik, K Weron, Complete description of all self-similar models driven by Lévy stable noise. *Phys. Rev. E* **71**, 016113 (2005).
- AV Weigel, B Simon, MM Tamkun, D Krapf, Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proc. Natl. Acad. Sci.* **108**, 6438–6443 (2011).
- T Akimoto, E Yamamoto, K Yasuoka, Y Hirano, M Yasui, Non-gaussian fluctuations resulting from power-law trapping in a lipid bilayer. *Phys. Rev. Lett.* **107**, 178103 (2011).
- SMA Tabei, et al., Intracellular transport of insulin granules is a subordinated random walk. *Proc. Natl. Acad. Sci.* **110**, 4911–4916 (2013).
- TA Welling, et al., Observation of undamped 3D Brownian motion of nanoparticles using liquid-cell scanning transmission electron microscopy. *Part. & Part. Syst. Charact.*, 2000003 (2020).
- MN Yesibolati, et al., Unhindered Brownian motion of individual nanoparticles in liquid phase scanning transmission electron microscopy. *Nano Lett.* (2020).
- JH Park, et al., Control of electron beam-induced Au nanocrystal growth kinetics through solution chemistry. *Nano Lett.* **15**, 5314–5320 (2015).
- SA Canepa, BT Sneed, H Sun, RR Unocic, K Mølhave, Influence of cetyltrimethylammonium bromide on gold nanocrystal formation studied by in situ liquid cell scanning transmission electron microscopy. *The J. Phys. Chem. C* **122**, 2350–2357 (2018).
- MR Hauwiler, et al., Dynamics of nanoscale dendrite formation in solution growth revealed through in situ liquid cell electron microscopy. *Nano Lett.* **18**, 6427–6433 (2018).

46. MR Hauwiler, JC Ondry, AP Alivisatos, Using graphene liquid cell transmission electron microscopy to study in situ nanocrystal etching. *JoVE*, e57665 (2018).
47. X Ye, J Chen, BT Dirroll, CB Murray, Tunable plasmonic coupling in self-assembled binary nanocrystal superlattices studied by correlated optical microspectrophotometry and electron microscopy. *Nano Lett.* **13**, 1291–1297 (2013).
48. LCG Rogers, Arbitrage with fractional Brownian motion. *Math. Finance* **7**, 95–105 (1997).



## Supporting Information Text

### Anomalous Diffusion Models

**Fractional Brownian Motion.** A FBM process  $x(t)$  is characterized by the following properties

- is a zero mean process  $\langle x(t) \rangle = 0$ .
- starts at  $x(0) = 0$ .
- has stationary increments

$$x(t) - x(0) \stackrel{d}{=} x(t + \Delta) - x(\Delta) \quad \forall \Delta, \quad [1]$$

where  $\stackrel{d}{=}$  denotes equality in distribution. A consequence is that the expectation of any function  $f$  of an increment is invariant to time translation of that increment; that is,

$$\langle f(x(t) - x(0)) \rangle = \langle f(x(t + \Delta) - x(\Delta)) \rangle. \quad [2]$$

Together with the previous property, this implies

$$\langle (x(t_1) - x(t_2))^2 \rangle = \langle (x(t_1 - t_2))^2 \rangle, \quad [3]$$

an identity which will be used shortly.

- has the probability density function (PDF) of the form (1)

$$P(x, t) = \frac{1}{\sqrt{4\pi D_H t^{2H}}} \exp\left(-\frac{x^2}{4D_H t^{2H}}\right). \quad [4]$$

with  $\langle x(t) \rangle = 0$ , and  $\langle x(t)^2 \rangle = 2D_H t^{2H}$ . Here,  $H$  is known as the Hurst exponent that is related to the anomalous diffusion exponent  $\alpha$  as  $H = \alpha/2$ . If  $0 < H < 1/2$  the process is subdiffusive, if  $H = 1/2$  the process is fully Brownian, and if  $1/2 < H < 1$  the process is super-diffusive. The second moment or the ensemble-averaged mean-squared displacement (e-MSD) of the FBM process is then

$$\langle x^2(t) \rangle = 2D_H t^{2H}. \quad [5]$$

With this definition and using a binomial expansion and using stationarity and zero mean properties of the last term on the second line ( $\langle x^2(t_1) - x^2(t_2) \rangle = \langle x^2(t_1 - t_2) \rangle = 2D_H(t_1 - t_2)^{2H}$ ), and finally using Eq. (5) for each term, the FBM process  $x(t)$  has a covariance\* of the form (2)

$$\begin{aligned} \langle x(t_1)x(t_2) \rangle &= \frac{1}{2} \langle x(t_1)^2 + x(t_2)^2 - (x(t_1) - x(t_2))^2 \rangle \\ &= \frac{1}{2} \langle x(t_1)^2 + x(t_2)^2 - (x(t_1 - t_2))^2 \rangle \\ &= D_H(t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H}). \end{aligned} \quad [6]$$

It can be concluded from the covariance of equation Eq. (6) that the FBM process is self-similar

$$x(\lambda t) \stackrel{d}{=} \lambda^H x(t). \quad [7]$$

Note that  $H$  is also known as the self-similarity parameter. The FBM process (of which ordinary Brownian motion can be considered a subset with  $H = 1/2$ ) is the only Gaussian process that is both self-similar and stationary.

The time evolution of  $x(t)$  can be assumed to have the general form

$$x(t) = \int_0^t dt' \xi(t'), \quad [8]$$

where  $\xi(t')$  is called fractional Gaussian noise. Equivalently, in differential form

$$\frac{d}{dt}x(t) = \xi(t). \quad [9]$$

\*Note  $\langle \cdot \rangle$  denotes expectation  $\mathbb{E}(\cdot)$  and since this is a zero-mean process that equals the covariance  $\mathbb{E}(x(t_1)x(t_2)) = \text{cov}(x(t_1), x(t_2))$

This implies that the time correlation of the fractional Gaussian noise can be obtained by differentiating equation Eq. (6) with respect to each of the time variables

$$\begin{aligned}
\langle \xi(t_1)\xi(t_2) \rangle &= \frac{d}{dt_1} \frac{d}{dt_2} \langle x(t_1)x(t_2) \rangle \\
&= -D_H \frac{d}{dt_1} \frac{d}{dt_2} |t_1 - t_2|^{2H} \\
&= D_H \frac{d}{dt_1} \left( 2H|t_1 - t_2|^{2H-1} \text{sgn}(t_1 - t_2) \right) \\
&= D_H \left( 2H(2H-1)|t_1 - t_2|^{2H-2} + 4H|t_1 - t_2|^{2H-1} \delta(t_1 - t_2) \right) \\
&= 2D_H H(2H-1)|t_1 - t_2|^{2H-2},
\end{aligned} \tag{10}$$

where  $\delta(x)$  is the Dirac delta function and  $\text{sgn}(x)$  is the sign function. Here we have used the properties

$$\frac{d}{dx} |x| = \text{sgn}(x), \tag{11}$$

$$\frac{d^2}{dx^2} |x| = \frac{d}{dx} \text{sgn}(x) = 2\delta(x), \tag{12}$$

$$\text{sgn}^2(x) = 1. \tag{13}$$

Finally, note that in the final line of equation Eq. (6) we have assumed that  $2H \neq 1$  to eliminate the second term in parentheses of the previous line. If instead we examine the case  $2H = 1$ , only the second term remains, leaving

$$\langle \xi(t_1)\xi(t_2) \rangle = 2D_H \delta(t_1 - t_2), \tag{14}$$

which is the expected delta-correlated noise characterizing Brownian motion.

**Discrete time FBM.** Here we reconsider the above analysis from the perspective of a discretized time variable, as will be made use of in the following section. From equation Eq. (6) the covariance for discrete-time increments of  $x_{t_2} - x_{t_1}$  and  $x_{s_2} - x_{s_1}$  is (3, 4)

$$\langle (x_{t_2} - x_{t_1})(x_{s_2} - x_{s_1}) \rangle = D_H((t_2 - s_1)^{2H} - (t_1 - s_1)^{2H} - (t_2 - s_2)^{2H} + (t_1 - s_2)^{2H}). \tag{15}$$

Hence, assuming that particle is at  $x = 0$  at time zero, the covariance of increments  $(x_{k+1} - x_k)$  and  $(x_1 - x_0)$  is

$$\langle (x_{k+1} - x_k)(x_1 - x_0) \rangle = D_H(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}). \tag{16}$$

Note that for  $H = 1/2$  (*i.e.*, Brownian case) these increments are non correlated and the process is not self-similar as we expect. The increments of the FBM process are also called fractional Gaussian noise  $\xi$ , where  $\xi(k+1) = x_{k+1} - x_k$  defined on increment of  $\partial t = 1$  ( $\frac{\partial x}{\partial t} = \xi(t)$ ). Therefore, Eq. (16) is indeed the covariance of the fractional Gaussian noise  $\langle \xi(k+1)\xi(1) \rangle$ . One can rewrite Eq. (16) by factoring the  $k^{2H}$  term

$$\langle \xi(k+1)\xi(1) \rangle = D_H k^{2H} f(1/k), \quad \text{for } k \geq 1 \tag{17}$$

where  $f(x) = (1-x)^{2H} + (1+x)^{2H} - 2$  (4). Using the Taylor expansion of  $f(x)$  at the origin ( $x = 1/k \rightarrow 0$ ) the covariance of the fractional Gaussian noise is

$$\gamma(k) = \langle \xi(k+1)\xi(1) \rangle = 2D_H H(2H-1)k^{2H-2}, \tag{18}$$

with  $k \in \{0, \dots, N-1\}$  and  $\gamma(0) = 1$ . This is true only if  $t_1 \neq t_2$  (4).

**Simulating a FBM process.** A FBM process can be simulated using a circulant matrix embedding algorithm and using fractional Gaussian noise  $\xi = (\xi_1, \xi_2, \dots, \xi_N)^T$  and its covariance matrix:

$$\mathbf{\Gamma} = \text{cov}(\xi) = \begin{pmatrix} 1 & \gamma(1) & \gamma(2) & \dots & \gamma(N-2) & \gamma(N-1) \\ \gamma(1) & 1 & \gamma(1) & \dots & \gamma(N-3) & \gamma(N-2) \\ \gamma(2) & \gamma(1) & 1 & \dots & \gamma(N-4) & \gamma(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(N-2) & \gamma(N-3) & \gamma(N-4) & \dots & 1 & \gamma(1) \\ \gamma(N-1) & \gamma(N-2) & \gamma(N-3) & \dots & \gamma(1) & 1 \end{pmatrix}.$$

In order to simulate a FBM process  $x(t)$ , we need to find the square root of the  $\mathbf{\Gamma}$  matrix. Finding square roots of this matrix is hard. Hence, a more convenient method often used is to embed this matrix  $\mathbf{\Gamma}$  in a larger circulant matrix called  $\mathbf{C}$  of size  $2M \times 2M$  with  $M = 2N$ :

$$\mathbf{C} = \begin{pmatrix} \boxed{\begin{matrix} 1 & \gamma(1) & \dots & \gamma(N-1) \\ \gamma(1) & 1 & \dots & \gamma(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \dots & 1 \end{matrix}} & 0 & \gamma(N-1) & \gamma(N) & \dots & \gamma(2) & \gamma(1) \\ \gamma(N-1) & 0 & \gamma(N-1) & \dots & \gamma(3) & \gamma(2) & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \dots & 1 & \gamma(1) & \gamma(2) & \gamma(3) & \dots & \gamma(N-1) & 0 \\ 0 & \gamma(N-1) & \dots & \gamma(1) & 1 & \gamma(1) & \gamma(2) & \dots & \gamma(N-2) & \gamma(N-1) \\ \gamma(N-1) & 0 & \dots & \gamma(2) & \gamma(1) & 1 & \gamma(1) & \dots & \gamma(N-3) & \gamma(N-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(1) & \gamma(2) & \dots & 0 & \gamma(N-1) & \gamma(N-2) & \gamma(N-3) & \dots & \gamma(1) & 1 \end{pmatrix},$$

where the red box indicates the  $\mathbf{\Gamma}$  matrix. Since the matrix  $\mathbf{C}$  is circulant, it can be decomposed into  $\mathbf{C} = \mathbf{F}\mathbf{\Lambda}\mathbf{F}^*$  using Fourier transform, where  $\mathbf{F}$  is a unitary matrix and  $\mathbf{\Lambda}$  is a diagonal matrix of eigenvalues of matrix  $\mathbf{C}$ .  $\mathbf{F}^*$  denotes the conjugate transpose of  $\mathbf{F}$  and  $\mathbf{F}\mathbf{F}^* = \mathbf{I}$ . Therefore,  $\mathbf{F}\mathbf{C}\mathbf{F}^* = \mathbf{\Lambda}$ . We can generate the matrix  $\mathbf{\Lambda}$  using the eigenvalues (*i.e.*, FFT coefficients of  $\mathbf{C}$ ) (4):

$$\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{2N-1}) \quad \lambda_m = \sum_{j=0}^{2N-1} C_j \exp 2\pi i \frac{jm}{2N}, \quad j, m = 0, \dots, 2N-1 \quad [19]$$

with  $C_j$  the  $(j+1)$ th elements of the first row if  $\mathbf{C}$  matrix,  $i = \sqrt{-1}$ , and  $\mathbf{F}$  defined as:

$$F(j, m) = \frac{1}{\sqrt{2N}} \exp -2\pi i \frac{jm}{2N}. \quad j, m = 0, \dots, 2N-1 \quad [20]$$

To find the square roots of matrix  $\mathbf{C}$ , we can write  $\mathbf{C} = \mathbf{S}\mathbf{S}^*$  with  $\mathbf{S} = \mathbf{F}\mathbf{\Lambda}^{1/2}\mathbf{F}^*$  and  $\mathbf{\Lambda}^{1/2} = \text{diag}(\lambda_0^{1/2}, \lambda_1^{1/2}, \dots, \lambda_{2N-1}^{1/2})$  (5). The last step to simulate a FBM process is to multiply matrix  $\mathbf{S}$  with a vector  $\mathbf{V}$  with i.i.d. standard normal elements and take the first  $N$  elements corresponding to the fractional Gaussian noise vector  $\boldsymbol{\xi}$ .

**Continuous Time Random Walk.** A continuous time random walk (CTRW) process is a class of anomalous diffusion with a combination of random walks in space and time. Consider a test particle diffusing with a CTRW behavior where  $x(t)$  denotes the position of the particle at time  $t$ . The particle will make a random jump of distance  $\Delta x_i = x(t_i) - x(t_{i-1})$  after a waiting time of  $\tau_i = t_i - t_{i-1}$  in its previous site. After the jump, the process is renewed. For a CTRW process, we assume (6)

- The spatial step length  $\Delta x_i$ ,  $i = 1, 2, \dots$  are i.i.d. random variables drawn from the PDF  $\lambda(\Delta x)$
- The waiting times  $\tau_i$ ,  $i = 1, 2, \dots$  are i.i.d. random variables drawn from the PDF  $\psi(\tau)$
- The waiting times  $\tau_i$ ,  $i = 1, 2, \dots$  and step lengths  $\Delta x_i$ ,  $i = 1, 2, \dots$  are independent

Therefore, the joint probability distribution function  $\varphi(\Delta x, \tau)$  (known as the jump PDF) can be written as  $\varphi(\Delta x, \tau) = \psi(\tau)\lambda(\Delta x)$  (6), where the distribution of the spatial jump and waiting times are (7):

$$\lambda(\Delta x) = \int_0^{\infty} dt \varphi(\Delta x, \tau), \quad [21]$$

$$\psi(\tau) = \int_{-\infty}^{+\infty} dx \varphi(\Delta x, \tau). \quad [22]$$

We will now focus only on the subdiffusive CTRW process which is more relevant to the anomalous diffusion of gold nanorods in the liquid cell environment. For a subdiffusive CTRW process, the waiting times  $\tau$  are drawn from a heavy-tailed power-law distribution with the asymptotic behavior

$$\lim_{\tau \rightarrow \infty} \psi(\tau) = \frac{\tau_0^\alpha}{\tau^{1+\alpha}}. \quad [23]$$

Here,  $\tau_0$  is a scaling factor with the dimension of time. The average waiting time in the subdiffusive case ( $\alpha < 1$ ) diverges; that is  $\langle \tau \rangle = \int_0^{\infty} \tau \psi(\tau) d\tau \rightarrow \infty$ . The power-law distributed waiting times can be thought of as a physical picture where tracer particles are continually caught in potential wells with various depths (8, 9). The spatial step lengths are assumed here only to have zero mean and finite variance.

As mentioned in the main text, for a Brownian particle, the ensemble-averaged MSD (e-MSD)  $\langle x^2(t) \rangle$  grows linearly in time. However, for a subdiffusive CTRW process of total duration  $T$ , the e-MSD is (10)

$$\langle x^2(T) \rangle \sim T^\alpha. \quad [24]$$

To obtain this form, we begin by considering the ensemble average of time averaged of MSD (et-MSD),  $\overline{\langle \delta x^2 \rangle}$ , over an ensemble of independent trajectories of duration  $T$

$$\overline{\langle \delta x^2(\Delta; T) \rangle} = \frac{1}{T - \Delta} \int_0^{T-\Delta} \langle (x(t + \Delta) - x(t))^2 \rangle dt. \quad [25]$$

The integrand can be expressed in terms of the variance of the jump length  $\langle \delta x^2 \rangle$  and the average number of jumps  $n(t, t + \Delta)$  in the time span of  $(t, t + \Delta)$  as (10, 11):

$$\begin{aligned} \langle (x(t + \Delta) - x(t))^2 \rangle &= \langle \delta x^2 \rangle \langle n(t, t + \Delta) \rangle \\ &= \langle \delta x^2 \rangle [\langle n(0, t + \Delta) \rangle - \langle n(0, t) \rangle]. \end{aligned} \quad [26]$$

For a subdiffusive CTRW process, the average number of jumps for a specified time interval corresponds to a fractional Poisson process with  $\langle n(0, t) \rangle \sim t^\alpha$ . Therefore,

$$\begin{aligned} \overline{\langle \delta x^2(\Delta; T) \rangle} &\sim \frac{\langle \delta x^2 \rangle}{T - \Delta} \int_0^{T-\Delta} dt' [\langle n(0, t' + \Delta) \rangle - \langle n(0, t') \rangle] \\ &= \frac{\langle \delta x^2 \rangle}{T - \Delta} \int_0^{T-\Delta} dt' [(t' + \Delta)^\alpha - t'^\alpha] \\ &= \langle \delta x^2 \rangle \frac{T^{1+\alpha} - \Delta^{1+\alpha} - (T - \Delta)^{1+\alpha}}{(1 + \alpha)(T - \Delta)}. \end{aligned} \quad [27]$$

In the limit  $\Delta \ll T$ :

$$\overline{\langle \delta x^2(\Delta; T) \rangle} \sim D_\alpha \frac{\Delta}{T^{1-\alpha}}, \quad [28]$$

which shows a linear dependence on time delay  $\Delta$  despite the nonlinear anomalous diffusive behavior with the measurement time  $T$ . The e-MSD of eq. Eq. (24) corresponds to the limit  $\Delta \rightarrow T$ . The fact that the measurement time  $T$  shows up in the eq. (28) shows the *aging* behavior of the subdiffusive CTRW process. This suggests that as the CTRW process goes on in time, the t-MSD becomes smaller, meaning that it is more likely that longer trapping times would happen, which stalls the progress of  $x(t)$  (1). Moreover, we observe a drastic difference between the subdiffusive CTRW and Brownian motion: that the t-MSD  $\overline{\delta x^2(\Delta; T)}$  and e-MSD  $\langle x^2(T) \rangle$  do not converge towards agreement even in the limit of infinite sampling, a condition known as *weak ergodicity breaking*

$$\langle x^2(T) \rangle \neq \lim_{T \rightarrow \infty} \overline{\delta x^2(\Delta; T)}. \quad [29]$$

This ergodicity-breaking nature of the CTRW process results in scatter in t-MSD  $\overline{\delta^2(\Delta)}$  vs. time delay  $\Delta$  curves.

### ***p*-Variation Test**

The *p*-variation test introduced in the Methods section generalizes the concept of the total variation  $V$ , in which the increments (*i.e.*, particle displacements) are summed over the entire trajectory

$$V[x(t)] = \lim_{n \rightarrow \infty} \sum_{j=1}^{2^n} |x(j/2^n) - x((j-1)/2^n)|. \quad [30]$$

Here we have expressed the total variation as a functional of the trajectory  $x(t)$  and rescaled the duration of this trajectory to the interval  $t \in [0, 1]$ . The total variation  $V[x(t)]$  measures the total length of the path traced out by  $x(t)$ . It is defined in the limit of  $n \rightarrow \infty$ . In this limit, the total variation of Brownian motion is infinite as will be shown momentarily. This is a simple example of the “coastline paradox” described by Benoit Mandelbrot in the context of self-similarity and fractal dimension: the total length of a continent’s coastline depends on the size of the ruler used to measure it and, in principle, can be infinite for an infinitesimal ruler (12).

The *p*-variation  $V_n^{(p)}(t)$  generalizes the concept of total variation by exponentiating each increment (see Fig. S4 for the increments size at each  $n$ ) by  $p$  before summing (13)

$$V_n^{(p)}(t) = \sum_{j=1}^{(2^n)t} |x(j/2^n) - x((j-1)/2^n)|^p. \quad [31]$$

Note that  $V_n^{(p)}(t)$  is defined for finite  $n$  and on any interval of the trajectory  $[0, t]$ .

We consider now the *p*-variation of fractional Brownian motion (FBM). The variance of FBM in Eq. (5) can be rewritten as

$$x(t + \delta) - x(t) \sim \delta^H, \quad [32]$$

where the symbol “ $\sim$ ” indicates expectational proportionality of the Euclidean norm. Together with the stationarity property of FBM, this allows equation Eq. (31) to be evaluated to

$$V_n^{(p)}(t) \sim \sum_{j=1}^{(2^n)t} (2^{-n})^{pH} = t(2^n)^{1-pH}. \quad [33]$$

Thus, in the limit of  $n \rightarrow \infty$  the  $p$ -variation falls into three regimes depending on the choice of  $p$ .

$$\lim_{n \rightarrow \infty} V_n^{(p)}(t) = \begin{cases} +\infty & \text{if } p < H^{-1} \\ t & \text{if } p = H^{-1} \\ 0 & \text{if } p > H^{-1} \end{cases}. \quad [34]$$

Earlier it was stated that the total variation of Brownian motion is infinite. This can be identified with the first case above, in which  $H = 1/2$  for Brownian motion and  $p = 1$  for the total variation. We can also see that the quadratic variation ( $p = 2$ ) of Brownian motion is finite and proportional to  $t$ ; that is,  $\lim_{n \rightarrow \infty} V_n^{(2)}(t) \propto t$ .

### Waiting Time Distribution

Subdiffusive behavior in the context of a CTRW process arises as a consequence of a heavy-tailed waiting time distribution, characterized by the asymptotic behavior described in Equation Eq. (23). In Fig. S11 we have plotted the distribution of waiting times for one of the trajectories at a dose rate of  $49 \text{ e}^-/\text{\AA}^2\text{s}$ , counting the time required for displacements larger than a radial threshold, where displacements below this radius are considered immobile. This figure shows the waiting time distributions for radial thresholds of 20 and 100 nm. The choice of the cut-off radius has a significant effect in the power-law exponent of the waiting time distribution. For small values ( $\leq 20$  nm) the distribution has a power law tail of  $\sim -2.0$ , suggesting that the  $\alpha \rightarrow 1$  corresponding to a Brownian case. However, as discussed in the text, displacements smaller than the length of the nanorods ( $\leq 60$  nm) could also mean that the nanorods got trapped with the head or tail on the same pinning site. Therefore, we have also plotted the distribution for a radial threshold value of 100 nm. However, the number of data points are insufficient to draw any firm conclusions, but the apparent asymptotic behavior in the 100 nm case may be an indication of subdiffusive behavior.

### MotionNet (MoNet) Architecture, Training, and Inference

**Input.** For diffusion classification, MoNet is trained on 10,000 simulated trajectories from three classes of Brownian, subdiffusive FBM, and subdiffusive CTRW. The steps on how to sample trajectories from these processes have been discussed in the previous sections. For FBM, the range of  $\alpha$  considered was  $0.2 \leq \alpha \leq 0.96$ . For CTRW the range of  $\alpha$  considered was  $0.1 \leq \alpha \leq 0.99$ . For  $\alpha$  prediction for both FBM and CTRW processes, MoNet is trained on 3,000 simulated trajectories for each task.

For the task of classifying the trajectories into their diffusion class and predicting the  $\alpha$  exponent for CTRW process, the input to MoNet is the vector of discrete-time increments of the simulated trajectories. Given a batch of  $N$  simulated trajectories  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ , where  $\mathbf{x}_i = (x_{i,0}, x_{i,1}, \dots, x_{i,299})|_{i=1}^N$ , the vector of discrete-time increments is defined as  $d\mathbf{x}_i = (x_{i,1} - x_{i,0}, \dots, x_{i,299} - x_{i,298}) = (dx_{i,1}, dx_{i,2}, \dots, dx_{i,299})$ . It has been reported previously that for the  $\alpha$  prediction task (in case of FBM processes) learning the velocity autocorrelation of a trajectory is more effective than the trajectory increments (14). Hence, we followed the same procedure and used the velocity autocorrelation of the discrete-time increments vector as the input for the MoNet with autocorrelation defined as  $d\mathbf{x}_i * d\mathbf{x}_i^T$ , where  $*$  denotes convolution and  $d\mathbf{x}_i^T$  is the transpose of vector  $d\mathbf{x}_i$ .

**Architecture.** Fig. S4 and S5 show the architecture of MoNet, adapted from Granik *et. al.* (14) and modified based on the  $p$ -variation method introduced in the previous section. We use the same architecture universally regardless of the task (regression/classification). The architecture of MoNet comprises of 4 layers where the first layer consists of 6 convolutional sublayers ( $f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}$ ) that are applied on the input data in parallel. The first 5 convolutional sublayers are three layers deep with *relu* activation units ( $relu(\cdot) = \max(\cdot, 0)$  for rectification of the feature map), batch normalization (normalizing the responses across features map), and max pooling (finding the maximum over a local neighborhood). The number of filters applied in all of these sublayers are set to 32. After training, each of these filters capture a certain distinct pattern along the trajectory (*e.g.*, descending, ascending patterns). The diversity among the filters are typically ensured via random initialization of the filters and regularization techniques such as batch normalization and drop out. The filter sizes are  $k = 3, 4, 2, 10,$  and  $20$  respectively for the five convolutional sublayers to capture the local dynamics of trajectories in several spacial resolutions. The convolutional sublayers also differ in their dilation factor (*i.e.*, the number of steps that filters skip). Following  $p$ -variation we chose dilation factors that span the trajectory via steps of size  $2^n$ . The last convolutional sublayer,  $f_{16}$  augments the model using large filter sizes of length 20 without any dilation. The output of the convolutional sublayers are fed into two fully connected layers of size 512 and 128 ( $f_2$  and  $f_3$ , respectively). The final layer of MoNet ( $f_4$ ) is set based on the prediction task. For the anomalous classification task, the last layer is a dense layer of size 3 (corresponding to the three classes of diffusion) with a Softmax activation. Softmax function maps the output of the layer 3 after applying the

linear transformation  $\mathbf{g}(\mathbf{dx}_i; \mathbf{W}) = [g_1(\mathbf{dx}_i; \mathbf{W}), g_2(\mathbf{dx}_i; \mathbf{W}), g_3(\mathbf{dx}_i; \mathbf{W})]$ , where  $\mathbf{W}$  denotes all the parameters in MoNet, to the predicted probability of output classes  $\mathbf{P}$  defined as:

$$\mathbf{P}(\mathbf{dx}_i; \mathbf{W}) = \frac{e^{\mathbf{g}(\mathbf{dx}_i; \mathbf{W})}}{\sum_{c=1}^C e^{g_c(\mathbf{dx}_i; \mathbf{W})}}, \quad [35]$$

where  $C = 3$  is the number of classes and  $\mathbf{P}(\mathbf{dx}_i; \mathbf{W}) = [P_1(\mathbf{dx}_i; \mathbf{W}), P_2(\mathbf{dx}_i; \mathbf{W}), P_3(\mathbf{dx}_i; \mathbf{W})]$ .

For the regression task of finding the  $\alpha$  exponent, a dense layer of size 1 with a Sigmoid activation is used in the last layer to capture the output. Sigmoid function maps the output  $g(\mathbf{dx}_i; \mathbf{W})$  to a variable between 0 and 1 (*i.e.*, the predicted value of  $\alpha$ ), and is defined as:

$$S^p(\mathbf{dx}_i; \mathbf{W}) = \frac{1}{1 + e^{-g(\mathbf{dx}_i; \mathbf{W})}}. \quad [36]$$

The overall architecture of the neural net shown in fig. S4 can be written as  $\mathcal{F}(\mathbf{dx}_i; \mathbf{W}) = f_4 \circ f_3 \circ f_2 \circ f_1(\mathbf{dx}_i; \mathbf{W})$  where  $f_1 = [f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}]$  is the concatenation of the output of all the 6 convolutional sublayers applied in the first layer (Fig. S5).

**Loss Function.** For classification task, the loss function is a categorical cross-entropy loss function,  $\mathcal{L}$ , defined as:

$$\mathcal{L}(\mathbf{W}) = \frac{1}{N} \sum_{i=1}^N D_{KL}(\mathbf{Q}_i || \mathbf{P}(\mathbf{dx}_i; \mathbf{W})) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C q_{i,c} \log P_c(\mathbf{dx}_i; \mathbf{W}), \quad [37]$$

where  $\mathbf{Q}_i = [q_{i,1}, q_{i,2}, q_{i,3}]$  is the ground truth probability of each class for a trajectory  $\mathbf{x}_i$ . Note that  $q_{i,c}$  is 1 if the sample  $i$  is in class  $c$  and 0 otherwise.  $P_c(\mathbf{dx}_i; \mathbf{W})$  is the output predicted probability that sample  $i$  is in class  $c$ .  $D_{KL}$  is the Kullback-Leibler divergence between two distributions  $\mathbf{P}_i$  and  $\mathbf{Q}_i$ .

For  $\alpha$  prediction, the loss function is a mean squared error (MSE)  $\mathcal{L}$  defined as:

$$\mathcal{L}(\mathbf{W}) = \frac{1}{N} \sum_{i=1}^N (S_i - S_i^p)^2, \quad [38]$$

with  $S_i^p$ , the predicted value of  $\alpha$  by MoNet (the output) and  $S_i$ , the ground truth value of  $\alpha$  for sample  $i$ .

**Training.** All the parameters of the network including the filters in the first layer and the weight matrices in the following layers were trained by back-propagating the derivative of the loss function with respect to the parameters  $\mathbf{W}$  using a stochastic gradient descent (15). MoNet is trained using the ADAM optimizer with an adaptive learning rate that starts from  $10^{-5}$ .

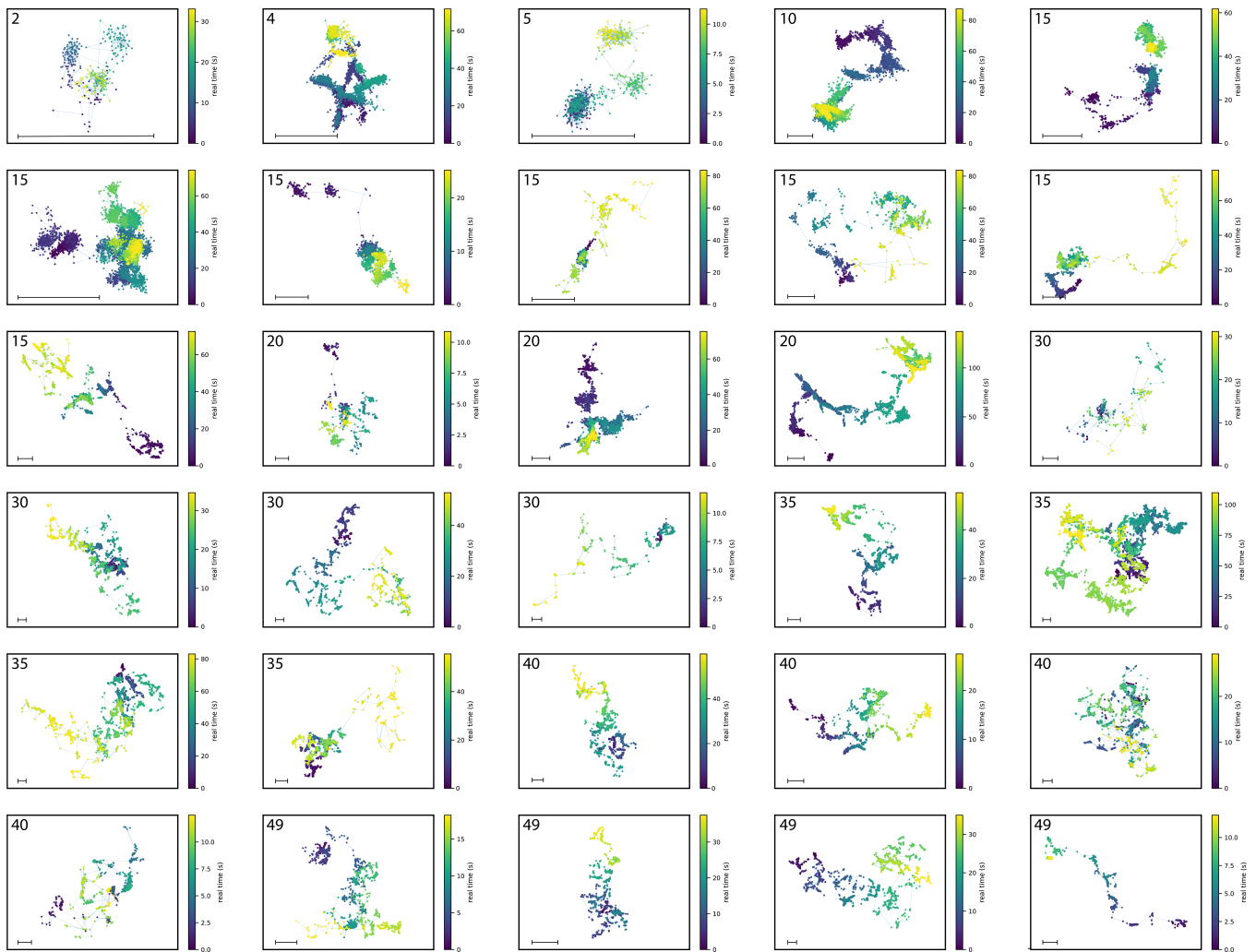
**Validation.** The validation accuracy and validation MSE are evaluated on a set of hold-out unseen simulated data with the same size as the training data (*i.e.*, 10,000 for classification and 3,000 for  $\alpha$  prediction).

**Inference.** For testing our experimental data (30 trajectories shown in Fig. S1), we treated  $x$  and  $y$  coordinates independently. For all trajectories  $(x_{i,1}, x_{i,2}, \dots, x_{i,T_i})_{i=1}^{30}$ , we tested each 300-frame intervals separately by dividing the trajectory into  $m = \lfloor T/300 \rfloor$  segments. The final results were then reported as the mean value of the output (probability in case of classification and  $\alpha$  value in case of  $\alpha$  prediction) averaged over all 300-frame segments and  $x$  and  $y$  coordinates. See Fig. S9 and S12 for the prediction outcomes for the 30 trajectories presented in Fig. S1.

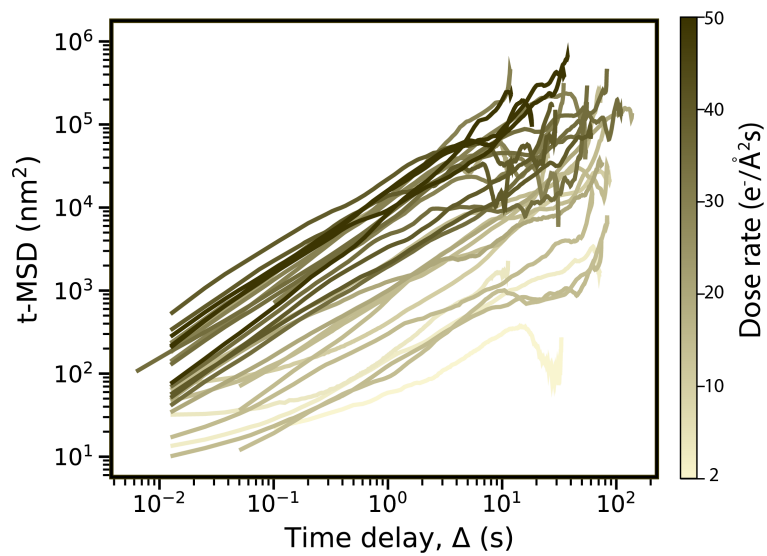
## Performance of MoNet in classification and $\alpha$ prediction

In order to show the effect of trajectory length on the performance of MoNet in both classification and  $\alpha$  prediction tasks, we have plotted Fig. S6 where we report validation accuracies and MES's averaged over 320 hold-out simulated trajectories. Fig. S6 shows that the accuracy increases and the MSE decreases, with increasing the trajectory length. However, the validation accuracy of MoNet for classification saturates around  $88.5\% \pm 2.3$  and validation MSE saturates over  $0.02 \pm 0.002$  for CTRW, and  $0.002 \pm 0.0002$  for FBM  $\alpha$  prediction, for trajectories longer than 300

In case of  $\alpha$  prediction, estimating  $\alpha$  based on a single trajectory and without having an ensemble average is a challenging task for CTRW processes. Therefore, as expected, the error associated with  $\alpha$  prediction for CTRW processes is higher than the case of FBM processes (Fig. S6a).

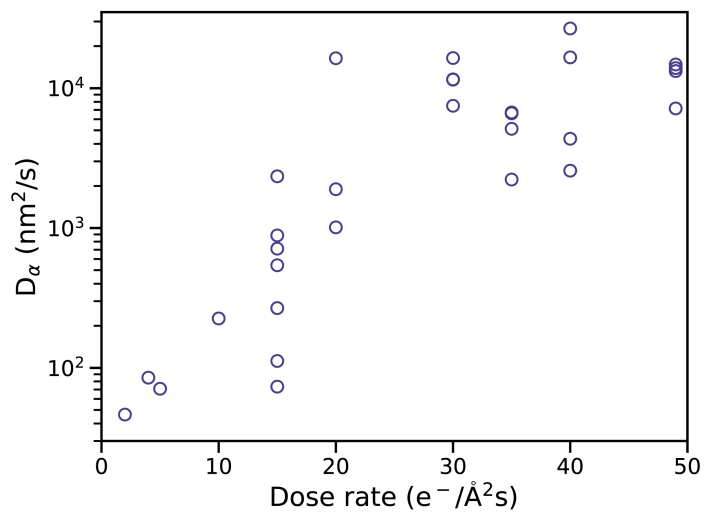


**Fig. S1.** Trajectories of 30 gold nanorods in a range of dose rates from 2 to 49  $e^-/\text{\AA}^2\text{s}$  over time (shown as color bar). Scale bars show 50 nm. Electron beam dose rate values are shown on the top left corner of each trajectory.

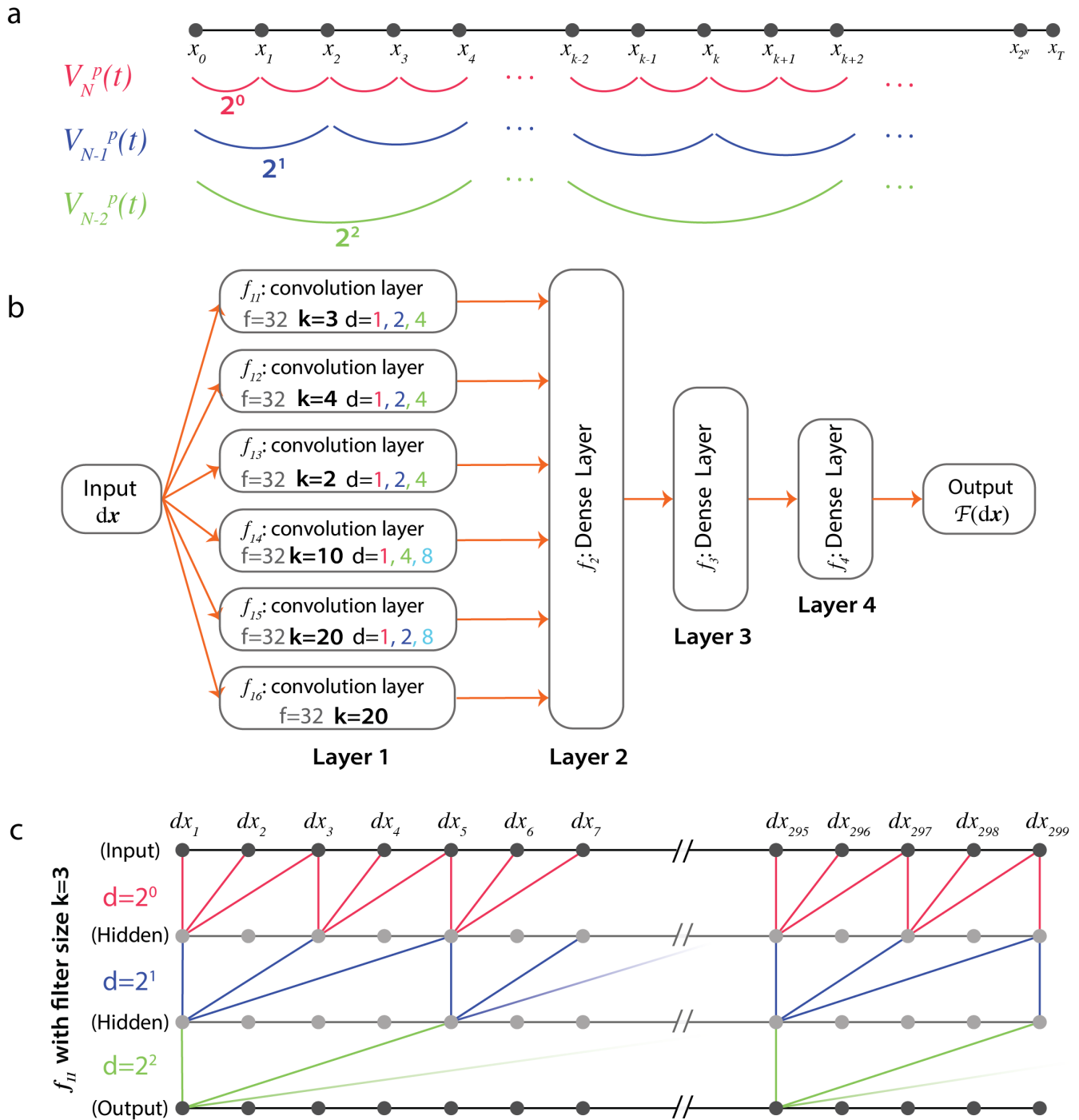


**Fig. S2.** Time-averaged mean squared displacement (t-MSD) vs. time delay,  $\Delta$ , for all trajectories of Fig. S1 in range of dose rates from 2 to 49 e<sup>-</sup>/Å<sup>2</sup>s.

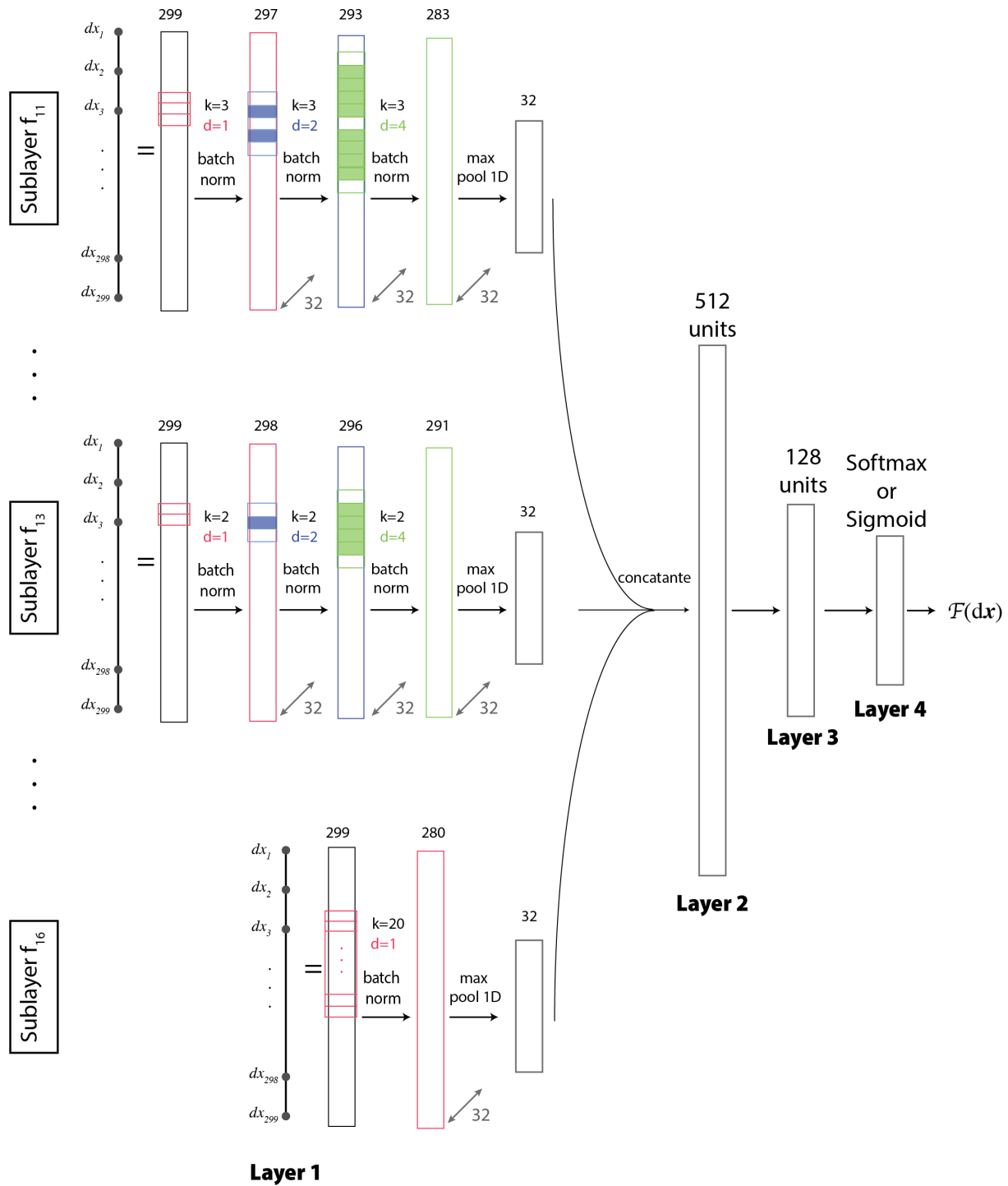




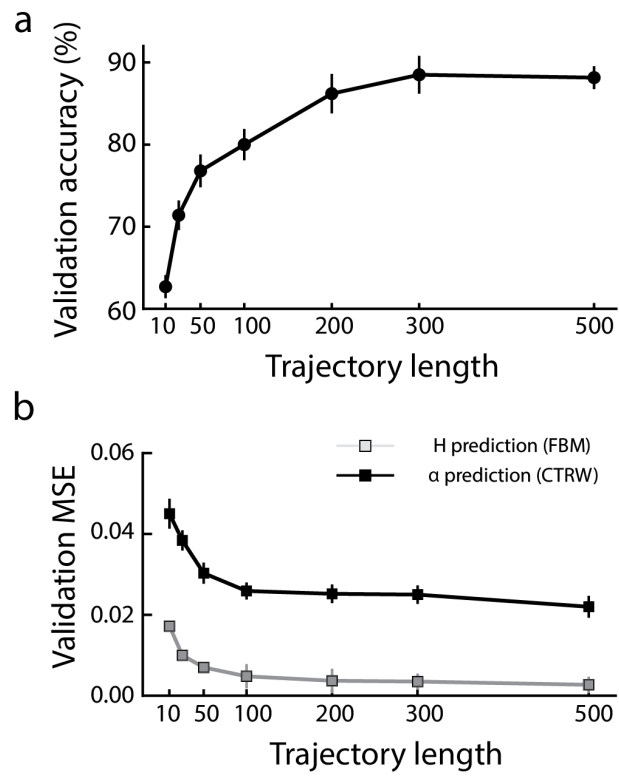
**Fig. S3.** Anomalous diffusion coefficients obtained from t-MSD curves of Fig. S2 fitted to a power law of  $D_\alpha \Delta^\alpha$  for time delays  $\Delta \leq 0.25$  s.



**Fig. S4.** a) Illustrative description of increments size for each given  $n$  in  $p$ -variation method,  $V_n^p(t)$  b) Schematic showing the Motion Net (MoNet) neural network architecture. c) An example of a dilated causal convolution sublayer  $f_{i1}$  used in MoNet with a filter size of  $k=3$  and dilation factors of  $d=2^0, 2^1, 2^2$  similar to the  $p$ -variation concept shown of panel (a).



**Fig. S5.** Architecture of the 4-layered MoNet model with 3 example convolutional sublayers (out of the 6 sublayers used in the first layer) with filter sizes of 3, 2, and 20. The input of the network is the increments of a trajectory,  $dx = (dx_1, dx_2, \dots, dx_{299})$ . In the first layer ( $f_{11}, \dots, f_{16}$ ) this input is convolved with 32 different filters of different sizes and different dilation factors (see Fig. S4). Here, we showed the example for filters of sizes:  $k = 3$  with dilations of  $d = 1$  (red), 2 (blue), and 4 (green),  $k = 2$  with dilations of  $d = 1$  (red), 2 (blue), and 4 (green), and  $k = 20$  with dilation of 1 (red). The output of each dilation is normalized using batch normalization. The resulting feature tensors are then pooled (maximum row within tensor), resulting in  $32 \times 1$  vector. Similar operation is performed on all sublayers of the first layer shown in Fig. S4. The result of all convolutional sublayers  $f_{11}, f_{12}, f_{13}, f_{14}, f_{15},$  and  $f_{16}$  were then concatenated into a vector of size  $192 \times 1$  which is the input for the second layer. The second and third layers are dense fully connected layers, taking features from the first layer. The final layer is chosen based on the prediction task; dense layer of size 3 with softmax activation function for the classification and dense layer of size 1 with sigmoid activation function for the  $\alpha$  prediction.



**Fig. S6.** a) Validation accuracy of MoNet in classifying the diffusion behavior on simulated test data. b) Validation mean squared error (MSE) associated with the  $\alpha$  prediction for CTRW (gray) and FBM (black) processes. Error bars indicate standard deviation.

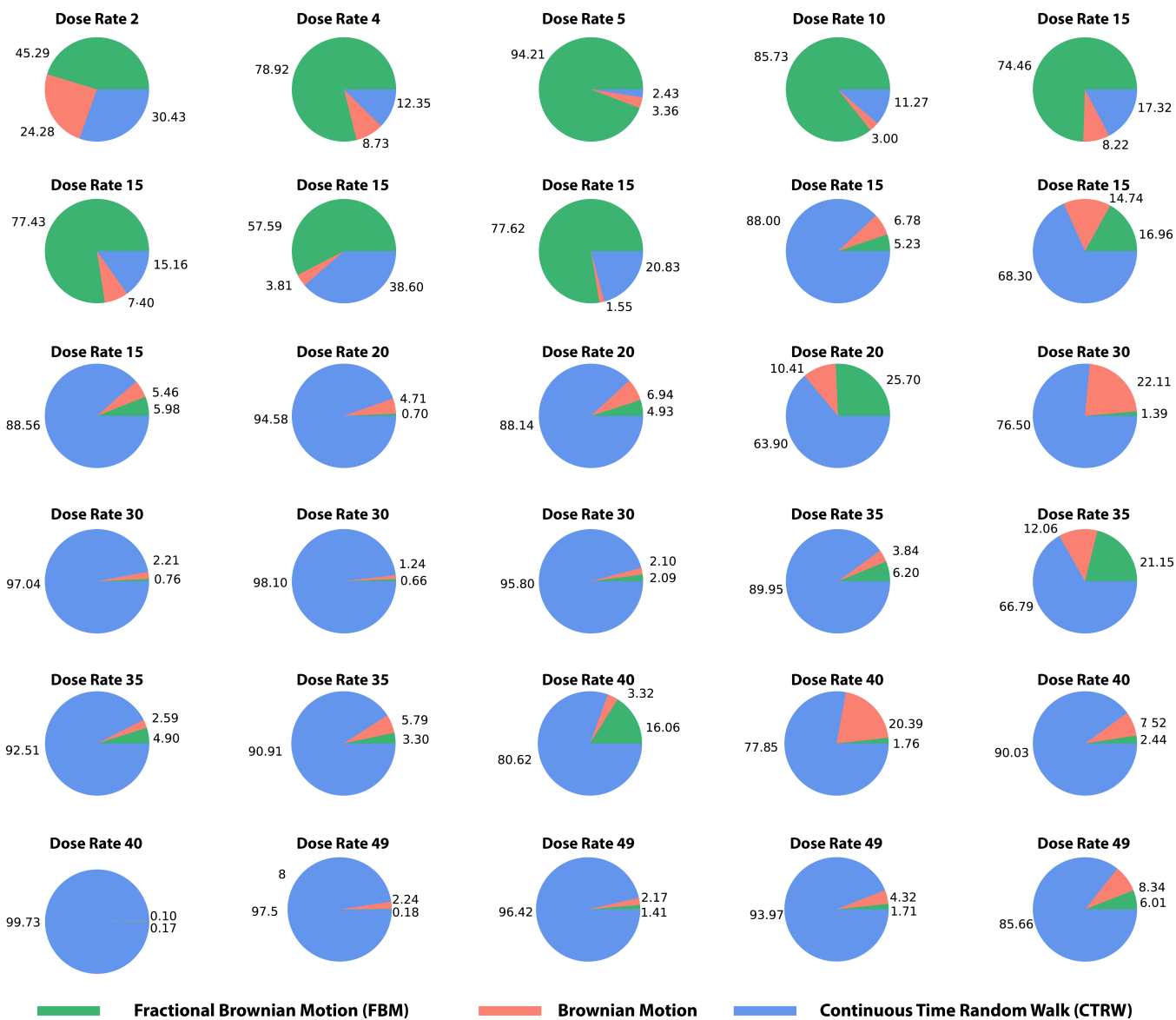


Fig. S7. Pie charts showing the predicted probability of three classes of diffusion by MoNet for trajectories shown in Fig. S1.

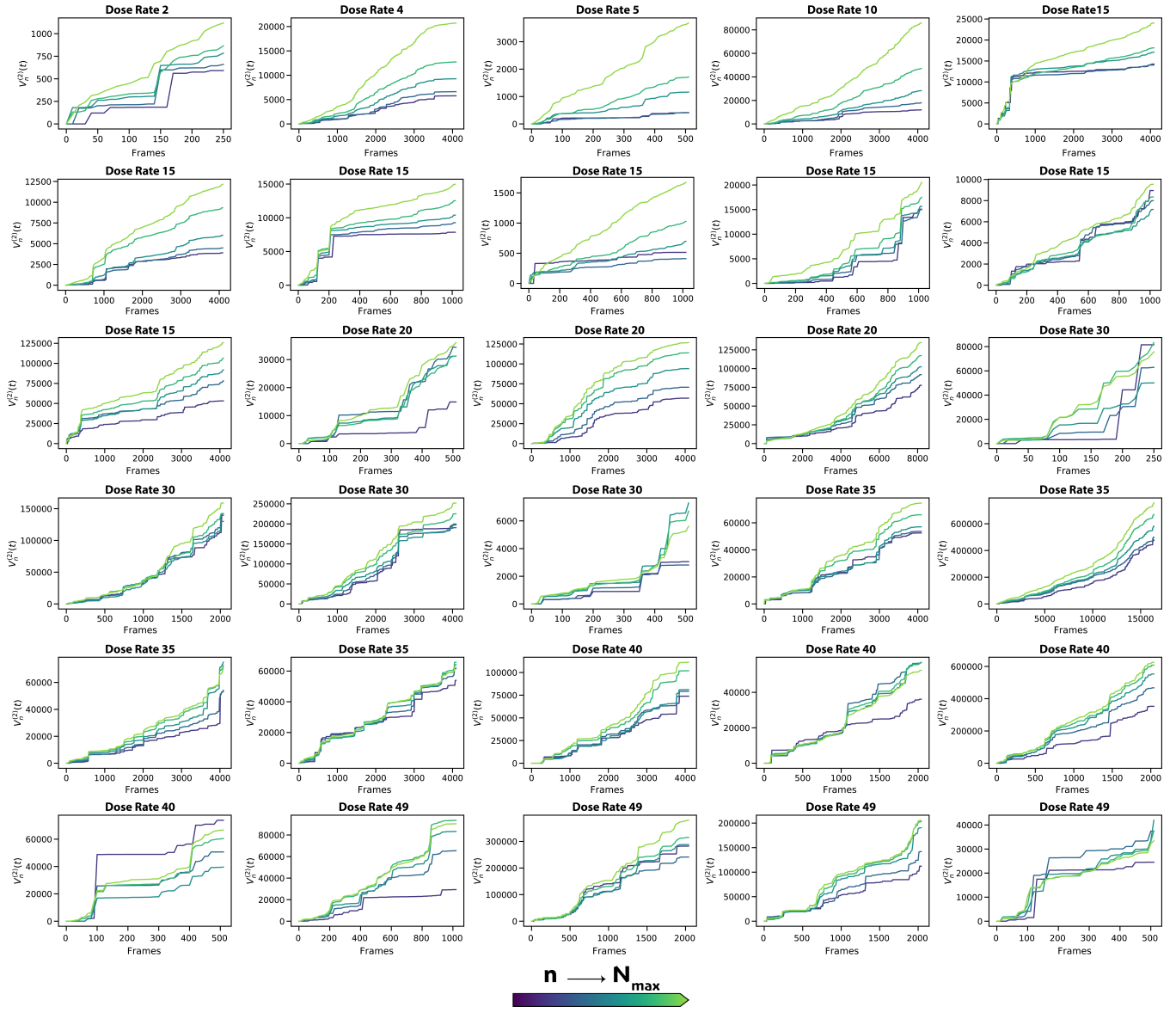


Fig. S8. Quadratic variation test,  $V_n^2(t)$ , results for trajectories shown in Fig. S1.

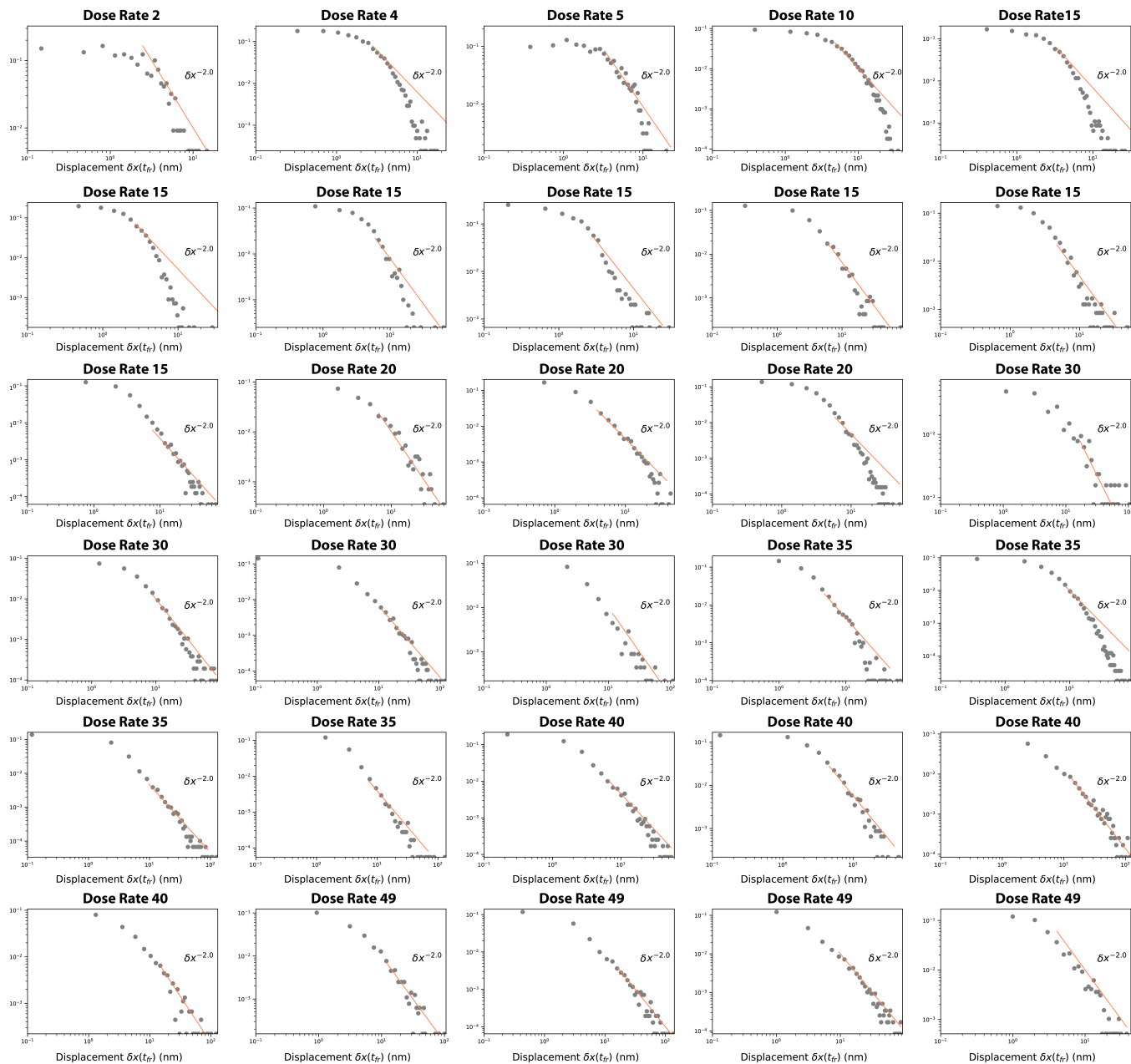
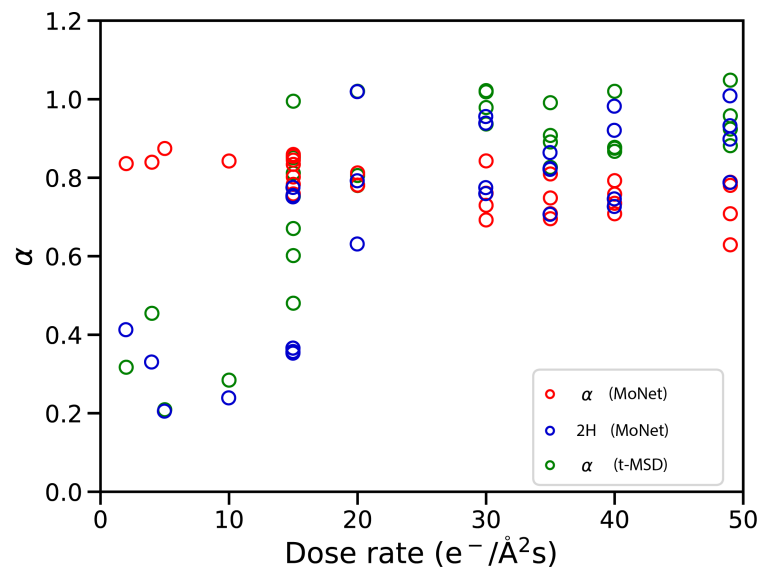
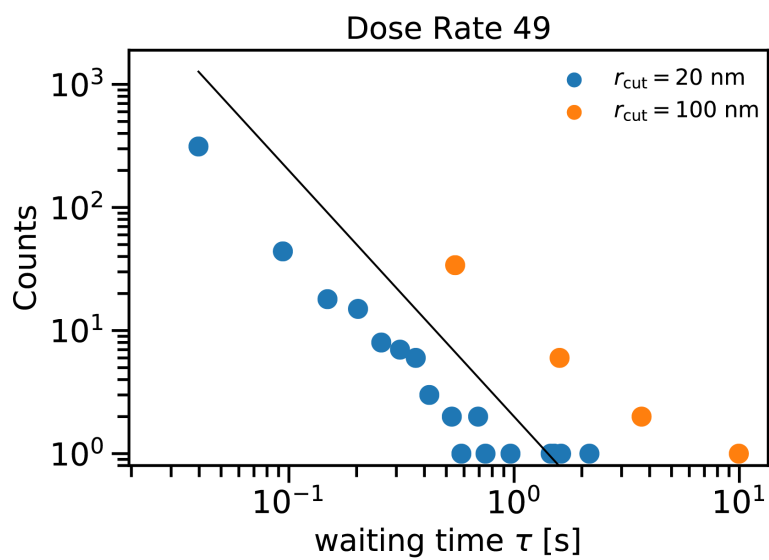


Fig. S9. Distribution of displacements,  $\delta x(t_{fr})$  (where  $t_{fr}$  is equal to  $1/\text{frame rate}$ ) for the trajectories shown in Fig. S1 with a power-law tail fit with an exponent of  $-2.0$ .

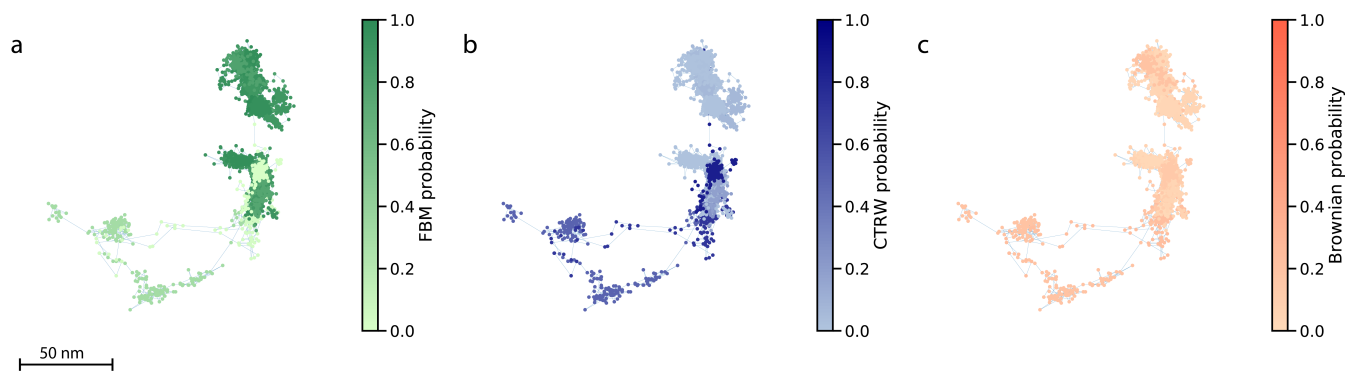


**Fig. S10.** Anomalous exponent  $\alpha$  predicted by MoNet for CTRW processes and FBM processes ( $\alpha = 2H$ , where  $H$  is the Hurst exponent of FBM processes) plotted as a function of dose rate and compared to the  $\alpha$  values estimated from the t-MSD curves of Fig. S2 for time delays  $\Delta \leq 0.25$  s.





**Fig. S11.** Distribution of waiting times,  $\tau$ , for an example trajectory at a dose rate of  $49 \text{ e}^-/\text{\AA}^2\text{s}$  for two different cut-off radius's of 20 and 100 nm. The black line indicates a power-law tail fit with an exponent of  $-2$ .



**Fig. S12.** Predicted probability of a) FBM, B) CTRW, and C) Brownian classes across the trajectory length captured at the electron beam dose rate of  $10 \text{ e}^-/\text{\AA}^2\text{s}$ .

## References

1. R Metzler, JH Jeon, AG Cherstvy, E Barkai, Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. Phys. Chem. Chem. Phys. **16**, 24128–24164 (2014).
2. J Picard, Representation formulae for the fractional brownian motion in Séminaire de probabilités XLIII. (Springer), pp. 3–70 (2011).
3. T Kaarakka, P Salminen, On fractional ornstein-uhlenbeck processes. Commun. on stochastic analysis **5**, 8 (2011).
4. T Dieker, Ph.D. thesis (Masters Thesis, Department of Mathematical Sciences, University of Twente) (2004).
5. O Banna, Y Mishura, K Ralchenko, S Shklyar, Fractional Brownian Motion: Approximations and Projections. (John Wiley & Sons), (2019).
6. G Germano, M Politi, E Scalas, RL Schilling, Stochastic calculus for uncoupled continuous-time random walks. Phys. Rev. E **79**, 066102 (2009).
7. R Metzler, J Klafter, The random walk’s guide to anomalous diffusion: a fractional dynamics approach. Phys. reports **339**, 1–77 (2000).
8. E Barkai, Y Garini, R Metzler, of single molecules in living cells. Phys. Today **65**, 29 (2012).
9. Y Meroz, IM Sokolov, J Klafter, Test for determining a subdiffusive model in ergodic systems from single trajectories. Phys. review letters **110**, 090601 (2013).
10. Y He, S Burov, R Metzler, E Barkai, Random time-scale invariant diffusion and transport coefficients. Phys. Rev. Lett. **101**, 058101 (2008).
11. S Burov, JH Jeon, R Metzler, E Barkai, Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking. Phys. Chem. Chem. Phys. **13**, 1800–1812 (2011).
12. Mandelbrot, B., How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. Science **156**, 636–638 (1967).
13. LC Rogers, Arbitrage with fractional brownian motion. Math. Finance **7**, 95–105 (1997).
14. N Granik, et al., Single-particle diffusion characterization by deep learning. Biophys. journal **117**, 185–192 (2019).
15. MD Zeiler, R Fergus, Visualizing and understanding convolutional networks in European conference on computer vision. (Springer), pp. 818–833 (2014).