

Supplementary Information for

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

Vida Jamali, Cory Hargus, Assaf Ben-Moshe, Amirali Aghazadeh, Hyun Dong Ha, Kranthi K. Mandadapu, and A. Paul Alivisatos

Corresponding Author.

E-mail: paul.alivisatos@berkeley.edu

This PDF file includes:

- Supplementary text
- Figs. S1 to S13
- Legends for Movies S1 to S5
- SI References

Other supplementary materials for this manuscript include the following:

- Movies S1 to S5

Supporting Information Text

1. 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 α 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 ξ , 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 $M = 2N$:

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

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 of \mathbf{C} matrix, $i = \sqrt{-1}$, and \mathbf{F} is 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 Δx_i , $i = 1, 2, \dots$ are i.i.d. random variables drawn from the PDF $\lambda(\Delta x)$
- The waiting times τ_i , $i = 1, 2, \dots$ are i.i.d. random variables drawn from the PDF $\psi(\tau)$
- The waiting times τ_i , $i = 1, 2, \dots$ and step lengths Δx_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 τ 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, τ_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 Δ despite the nonlinear anomalous diffusive behavior with the measurement time 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 Δ curves.

2. 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. S12 we have plotted the distribution of waiting times for one of the trajectories, collected at a dose rate of $49 \text{ e}^-/\text{\AA}^2\text{s}$, by 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 (≤ 20 nm) the distribution has a power-law tail of ~ -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 (≤ 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.

3. Statistical Analysis

Two-sided t-test and p-value Calculation. Two-sided t-test is used here to test for the null hypothesis that two independent classes of diffusion have identical probability. The p-values were calculated using `ttest_ind_from_stats` package from SciPy, which considers the mean, number of samples, and the standard deviation of predicted probability for FBM and CTRW classes of diffusion at any dose rate.

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 $x(t)$ and 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, $\lim_{n \rightarrow \infty} V_n^{(2)}(t) \propto t$.

4. 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 α considered was $0.2 \leq \alpha \leq 0.96$. For CTRW the range of α considered was $0.1 \leq \alpha \leq 0.99$. For α 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 α 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 $\mathbf{dx}_i = (x_{i,1} - x_{i,0}, \dots, x_{i,299} - x_{i,298}) = (dx_{i,1}, dx_{i,2}, \dots, dx_{i,299})$. It has been reported previously that for the α 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 $\mathbf{dx}_i * \mathbf{dx}_i^T$, where $*$ denotes convolution and \mathbf{dx}_i^T is the transpose of vector \mathbf{dx}_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 α 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 α), 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 α 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 α by MoNet (the output) and S_i , the ground truth value of α 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 α prediction).

Inference. For testing our experimental data (40 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 α value in case of α prediction) averaged over all 300-frame segments and x and y coordinates. See Fig. S10 and S13 for the prediction outcomes for the 40 trajectories presented in Fig. S1.

5. Performance of MoNet in classification and α prediction

In order to show the effect of trajectory length on the performance of MoNet in both classification and α prediction tasks, we have plotted Fig. S6 where we report validation accuracies and MSE'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 ± 0.002 for CTRW, and 0.002 ± 0.0002 for FBM α prediction, for trajectories longer than 300.

In case of α prediction, estimating α 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 α prediction for CTRW processes is higher than the case of FBM processes (Fig. S6a).

6. Performance of MoNet in classifying the behavior of hybrid trajectories

In order to show the power of MoNet in classifying the behavior of hybrid trajectories and predicting their corresponding α exponent, we have simulated interleaved trajectories where the diffusion mechanism randomly switches from one class to another class in each segment (Fig. S13a). The t-MSD curve, Fig. S13, for this hybrid interleaved trajectory is sublinear with a slope of 0.76. The p -variation test result shown in Fig. S13c may suggest that the dominant behavior is CTRW as the quadratic variation curves seem to be independent of the time step size Δt . This is while the behavior of the trajectory is 50% FBM and 50% CTRW with random permutations. This particular example shows that both t-MSD and p -variation are limited when it comes to trajectories with a hybrid behavior. Fig. S13e shows the ground truth class of diffusion along with MoNet predictions for each segment of the trajectory. The overall behavior of the trajectory can be determined by averaging the predictions over all segments of the trajectory. The results shown in the bar plot of Fig. S13f shows that MoNet predicts that the behavior is 50% CTRW and 36% FBM confirming that MoNet outperforms canonical methods in classifying the behavior of interleaved trajectories. Additionally, Fig. S13d shows that the slope of the t-MSD curve overestimates the value of α for this hybrid trajectory (0.76 *vs.* 0.6), while MoNet's prediction (0.56 ± 0.15) is closer to the ground truth value.

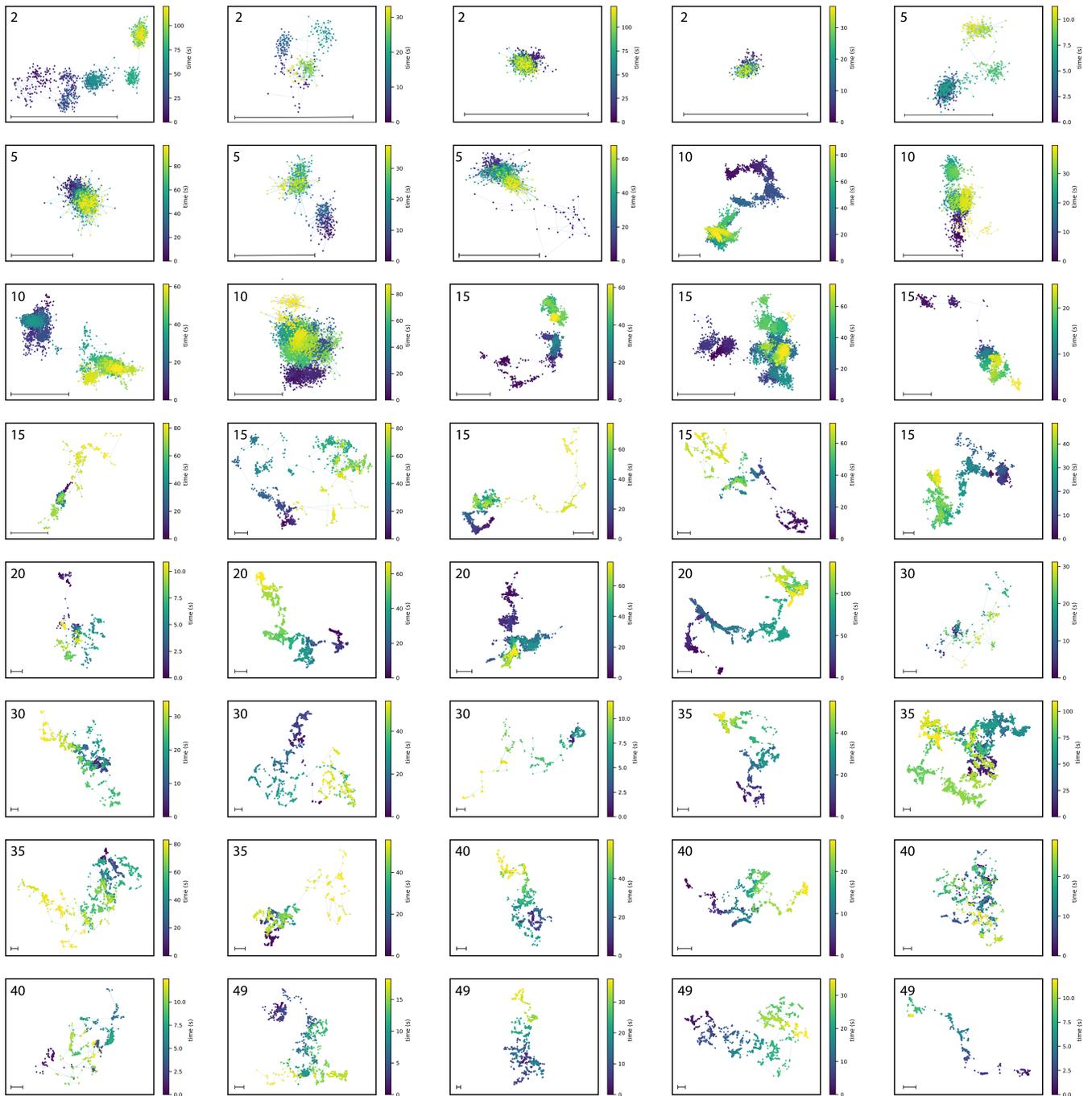


Fig. S1. Trajectories of 40 gold nanorods in a range of dose rates from 2 to 49 $e^-/\text{\AA}^2\text{s}$ over time (shown as color bar). Scale bars show 50 nm. Electron beam dose rate values are shown on the top left corner of each trajectory.

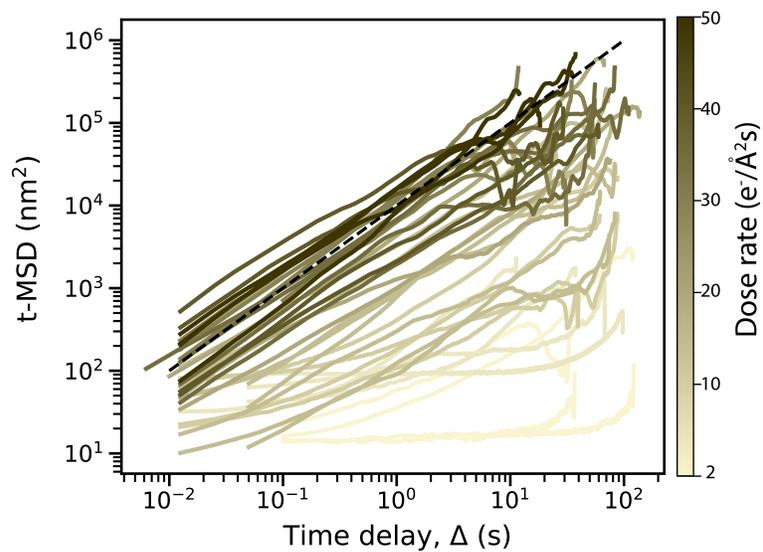


Fig. S2. Time-averaged mean squared displacement (t-MSD) vs. time delay, Δ , for all trajectories of Fig. S1 in range of dose rates from 2 to 49 e⁻/Å²s.

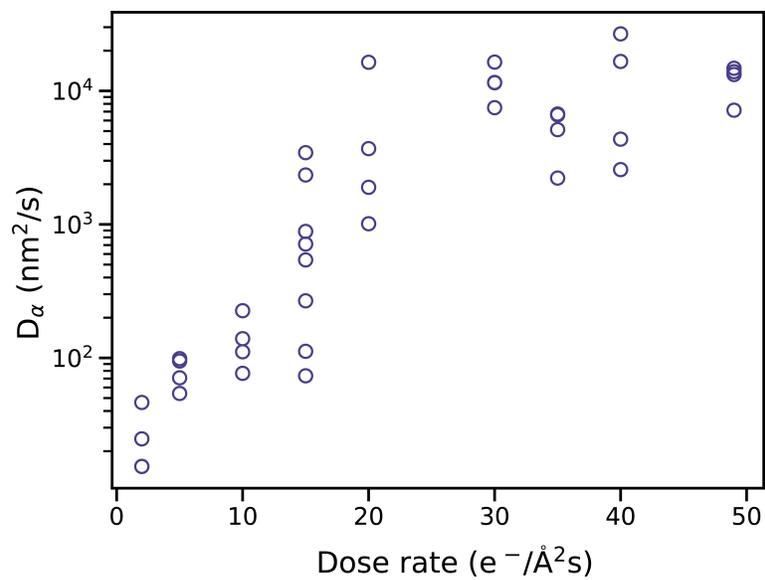


Fig. S3. Anomalous diffusion coefficients obtained from t-MSD curves of Fig. S2 fitted to a power law of $D_\alpha \Delta^\alpha$ for time delays $\Delta \leq 0.25$ s.

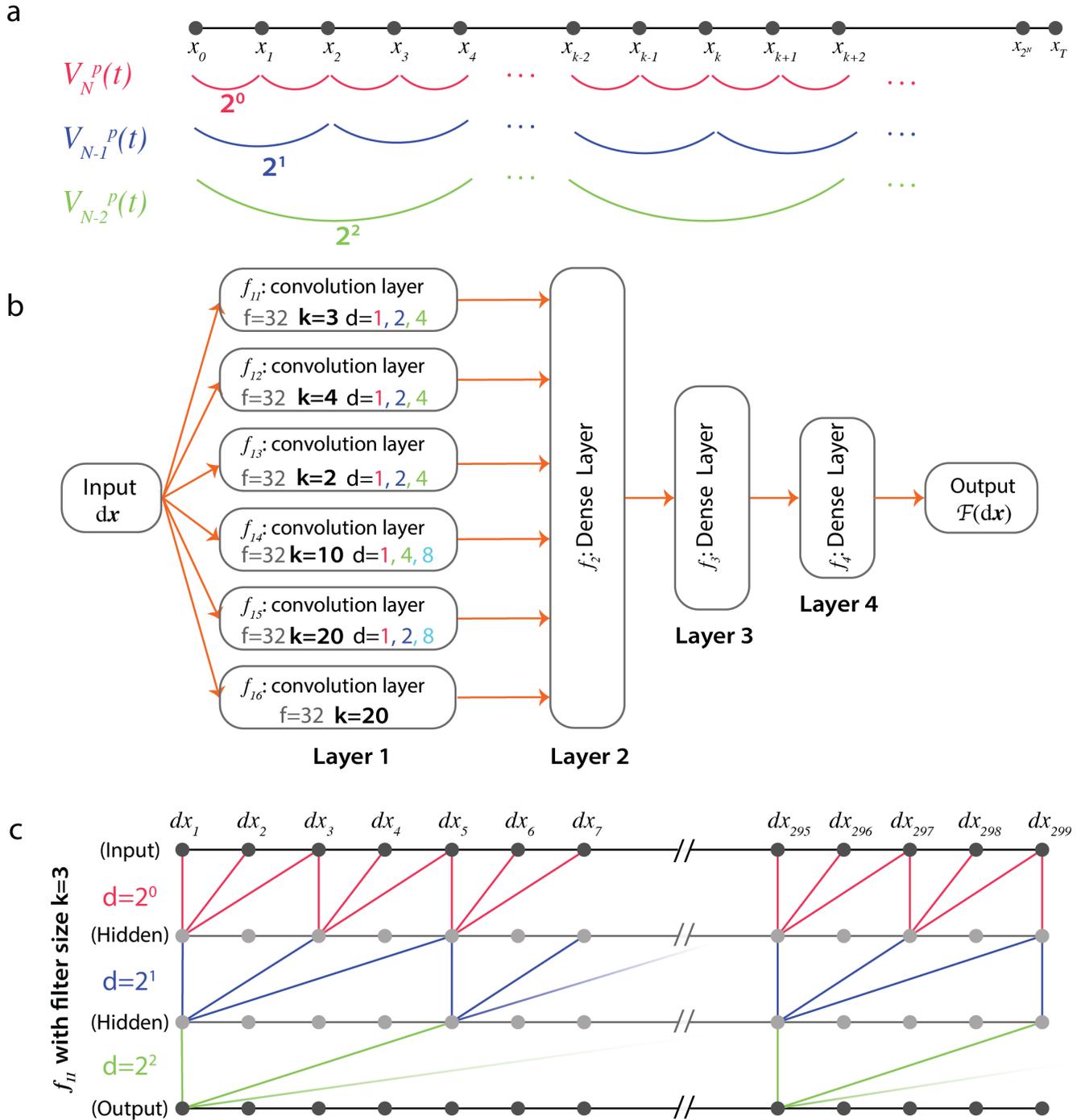


Fig. S4. a) Illustrative description of increments size for each given n in p -variation method, $V_n^P(t)$ b) Schematic showing the Motion Net (MoNet) neural network architecture. c) An example of a dilated causal convolution sublayer f_{11} used in MoNet with a filter size of $k=3$ and dilation factors of $d=2^0, 2^1, 2^2$ similar to the p -variation concept shown of panel (a).

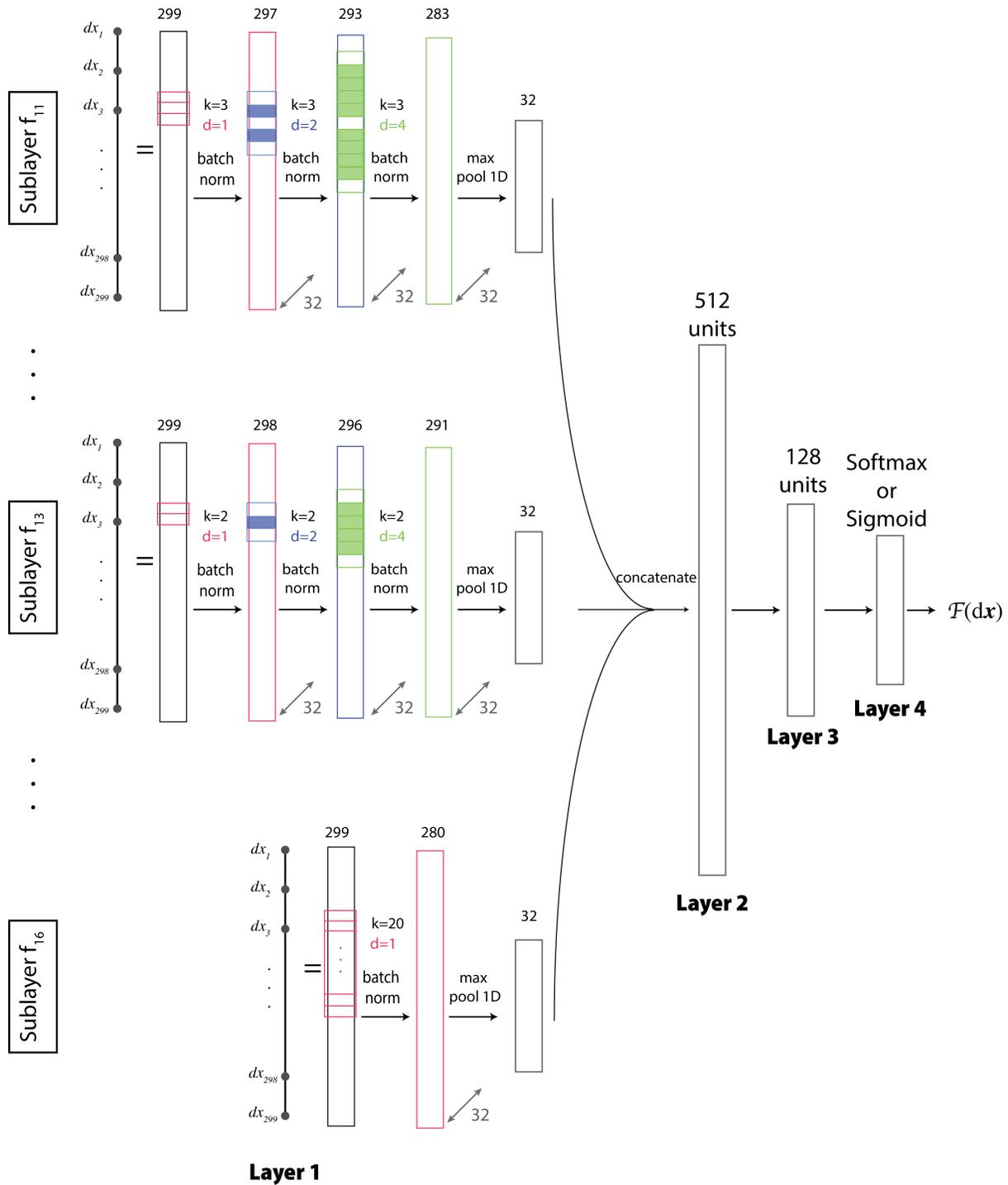


Fig. S5. Architecture of the 4-layered MoNet model with 3 example convolutional sublayers (out of the 6 sublayers used in the first layer) with filter sizes of 3, 2, and 20. The input of the network is the increments of a trajectory, $dx = (dx_1, dx_2, \dots, dx_{299})$. In the first layer (f_{11}, \dots, 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 32×1 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 192×1 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 α prediction.

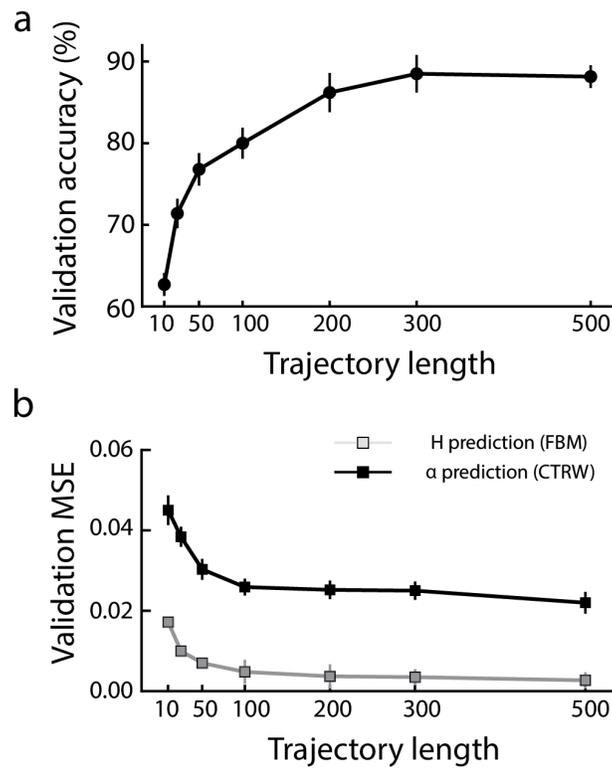


Fig. S6. a) Validation accuracy of MoNet in classifying the diffusion behavior on simulated test data. b) Validation mean squared error (MSE) associated with the α prediction for CTRW (gray) and FBM (black) processes. Error bars indicate standard deviation.



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

Dose Rate	p-value
2	5.3×10^{-7}
5	3.0×10^{-17}
10	9.0×10^{-58}
15	8.6×10^{-1}
20	3.2×10^{-25}
30	2.3×10^{-49}
35	3.9×10^{-58}
40	3.4×10^{-21}
49	1.0×10^{-47}

Fig. S8. p-values calculated using two-sided t-test at each dose rate. The results supports the presences of predominantly FBM behavior at low dose rates ($p < 5 \times 10^{-7}$), predominantly CTRW behavior at high dose rates ($p < 3 \times 10^{-21}$), and coexistence of both behaviors at the dose rate of $15 \text{ e}^-/\text{\AA}^2\text{s}$, where we cannot reject the null hypothesis that the probability of FBM and CTRW classes are equal ($p = 0.86$).

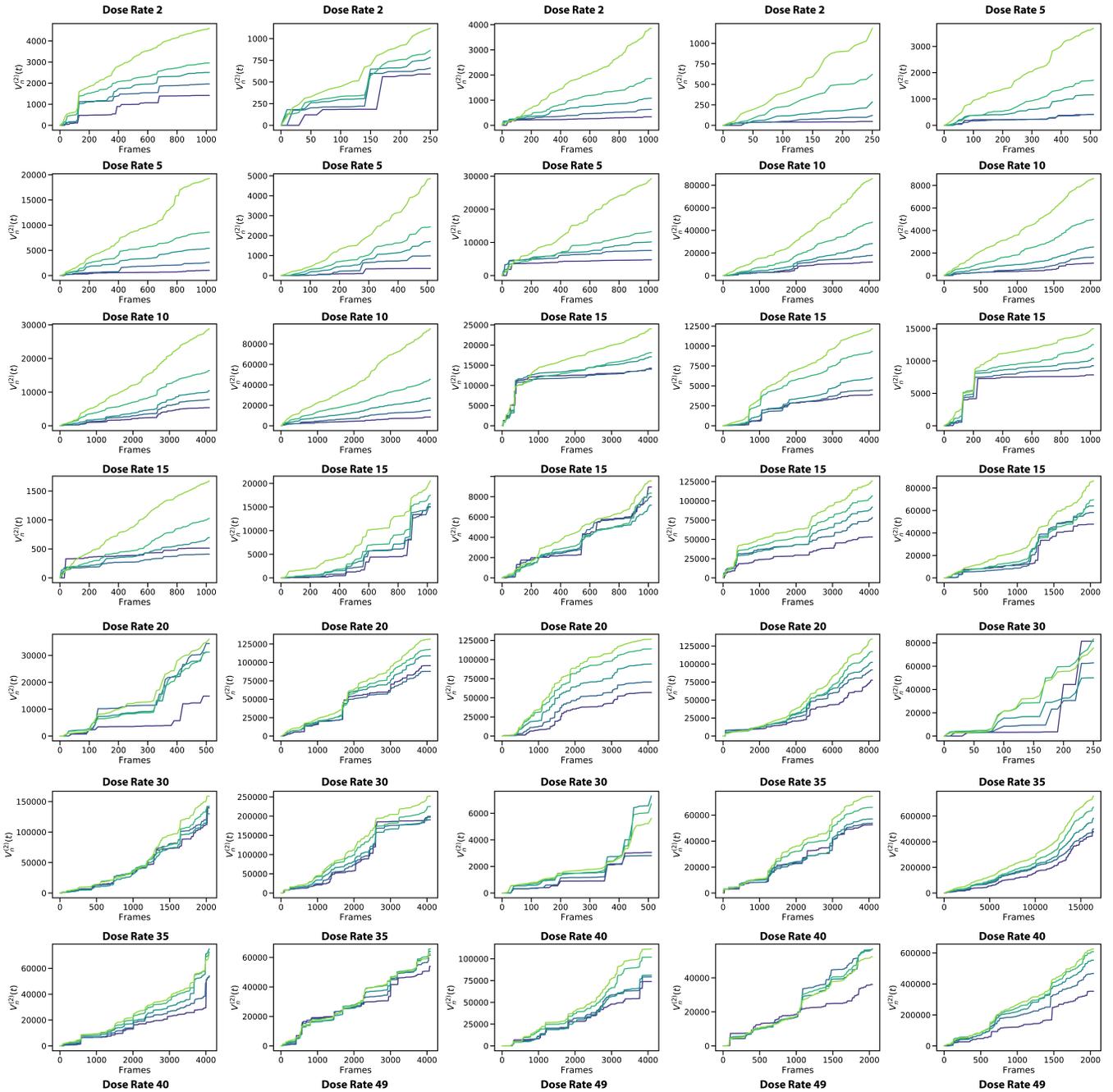


Fig. S9. Quadratic variation test, $V_n^{(2)}(t)$, results for trajectories shown in Fig. S1.

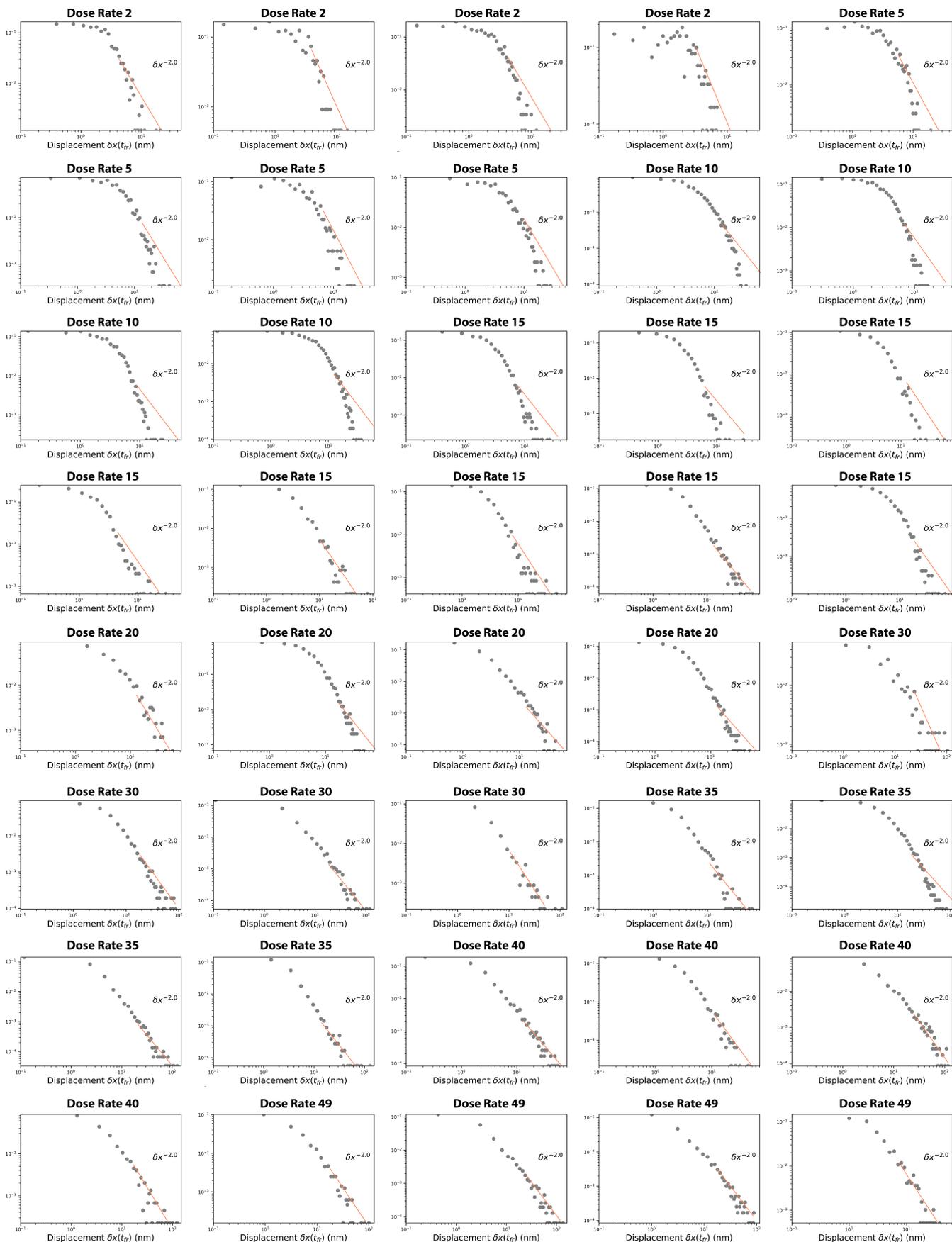


Fig. S10. Distribution of displacements, $\delta x(t_{Fr})$ (where t_{Fr} is equal to $1/\text{frame rate}$) for the trajectories shown in Fig. S1 with a power-law tail fit with an exponent of -2.0 .

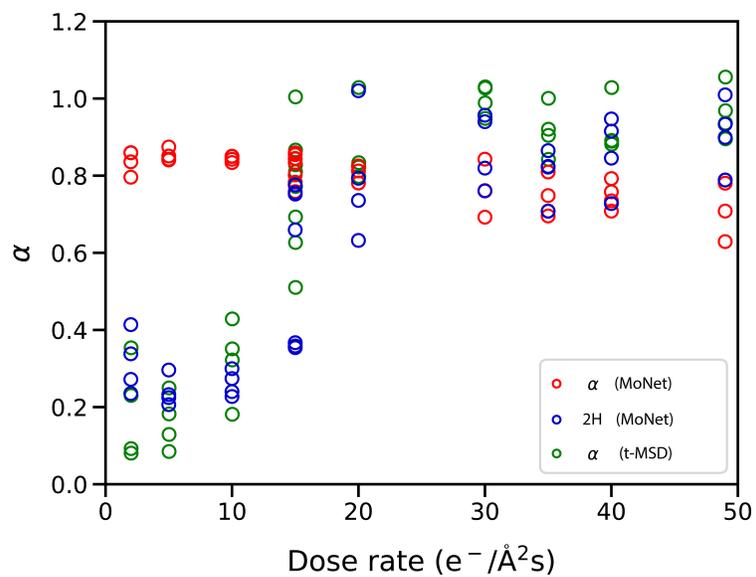


Fig. S11. Anomalous exponent α predicted by MoNet for CTRW processes and FBM processes ($\alpha = 2H$, where H is the Hurst exponent of FBM processes) plotted as a function of dose rate and compared to the α values estimated from the t-MSD curves of Fig. S2 for time delays $\Delta \leq 0.25$ s.

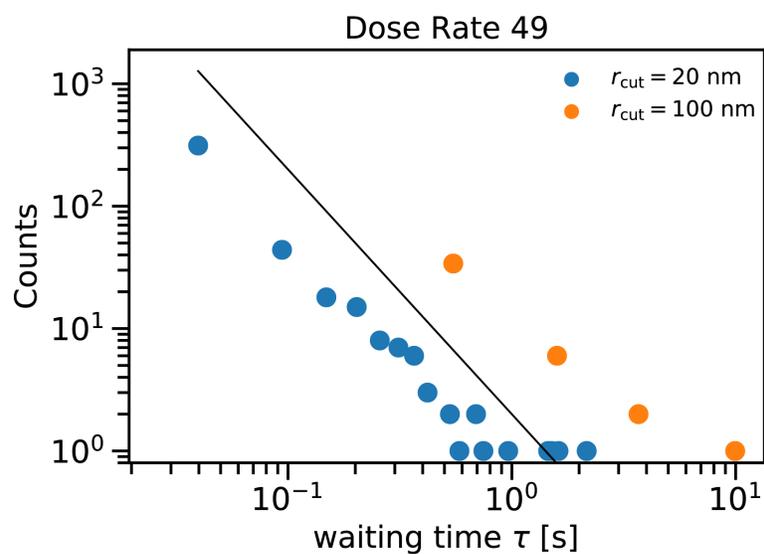


Fig. S12. Distribution of waiting times, τ , for an example trajectory at a dose rate of $49 \text{ e}^-/\text{\AA}^2\text{s}$ for two different cut-off radius values of 20 and 100 nm. The black line indicates a power-law tail fit with an exponent of -2 .

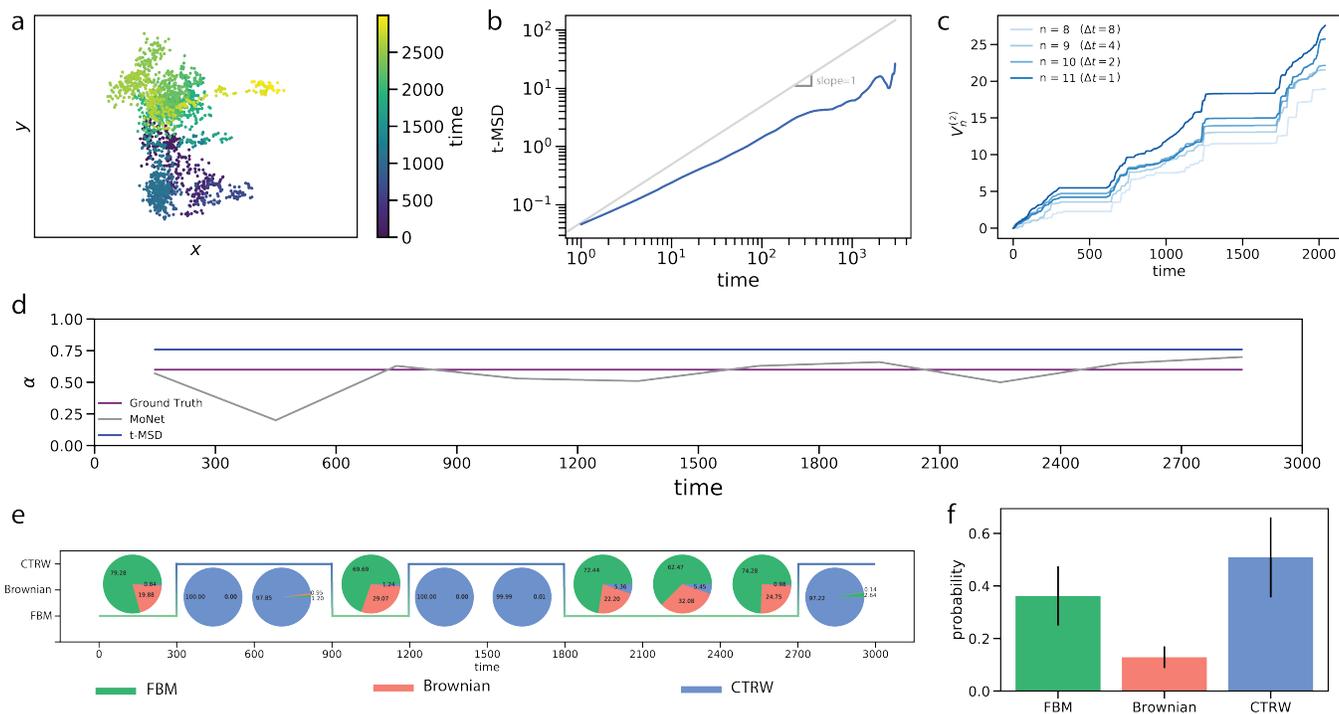


Fig. S13. a) A simulated interleaved trajectory with a hybrid behavior switching randomly between subdiffusive FBM and subdiffusive CTRW behavior ($\alpha = 0.6$). b) Corresponding t-MSD plot for the simulated hybrid trajectory in (a). The slope of the t-MSD curve is sublinear with an α exponent of 0.76. c) p -variation test results for the hybrid trajectory in (a). d) α values predicted by MoNet for each segment of the trajectory compared to the ground truth value (0.6) and over estimated value based on the slope of the t-MSD curve (0.76). e) Comparison of the ground truth class of diffusion with MoNet predictions for the dominant behavior. f) Bar plot showing the average behavior of the system along the entire length of the trajectory confirming the presence of both FBM and CTRW behavior in the simulated trajectory.

- Movie S1. Compressed and subsampled LCTEM video showing an AuNR motion at dose rate of $2 \text{ e}^-/\text{\AA}^2\text{s}$.
- Movie S2. Compressed and subsampled LCTEM video showing an AuNR motion at dose rate of $15 \text{ e}^-/\text{\AA}^2\text{s}$.
- Movie S3. Compressed and subsampled LCTEM video showing an AuNR motion at dose rate of $20 \text{ e}^-/\text{\AA}^2\text{s}$.
- Movie S4. Compressed and subsampled LCTEM video showing an AuNR motion at dose rate of $30 \text{ e}^-/\text{\AA}^2\text{s}$.
- Movie S5. Compressed and subsampled LCTEM video showing an AuNR motion at dose rate of $49 \text{ e}^-/\text{\AA}^2\text{s}$.

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).