

RESEARCH ARTICLE

Mean square displacement for a discrete centroid model of cell motion

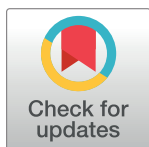
Mary Ellen Rosen¹, Christopher P. Grant, J. C. Dallon¹ ^{*}

Department of Mathematics, Brigham Young University, Provo, Utah, United States of America

 These authors contributed equally to this work.^{*} dallon@math.byu.edu

Abstract

The mean square displacement (MSD) is an important statistical measure on a stochastic process or a trajectory. In this paper we find an approximation to the mean square displacement for a model of cell motion. The model is a discrete-time jump process which approximates a force-based model for cell motion. In cell motion, the mean square displacement not only gives a measure of overall drift, but it is also an indicator of mode of transport. The key to finding the approximation is to find the mean square displacement for a subset of the state space and use it as an approximation for the entire state space. We give some intuition as to why this is an unexpectedly good approximation. A lower bound and upper bound for the mean square displacement are also given. We show that, although the upper bound is far from the computed mean square displacement, in rare cases the large displacements are approached.

 OPEN ACCESS

Citation: Rosen ME, Grant CP, Dallon JC (2021) Mean square displacement for a discrete centroid model of cell motion. PLoS ONE 16(12): e0261021. <https://doi.org/10.1371/journal.pone.0261021>

Editor: Alexandre Bovet, University of Oxford, UNITED KINGDOM

Received: April 22, 2021

Accepted: November 22, 2021

Published: December 20, 2021

Copyright: © 2021 Rosen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [S1 File](#).

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

One of the characteristics that distinguishes living things from non-living things is motility. On the cellular level, the motility or non-motility of different types of cells can be life-building (e.g., embryogenesis), life-saving (e.g., white blood cells) or life-threatening (e.g., metastasis). A thorough study of cell motility is needed to help understand the underlying mechanisms of motion in order to be able to inhibit or promote cell movement [1]. The mean square displacement (MSD) is often used to analyze particle motion and can be a first indicator of the mechanism of motion in cells. We give a brief introduction to the MSD to give a greater understanding of its origins and applications, including cell motion. It will later be applied to a math model of cell motion.

The MSD is a statistical measure of the average distance a particle travels over time. It can be thought of as a measure of overall drift. For instance, if a particle has a lot of motion within a small radius, its displacement over time may not be a good measure of overall motion, whereas the MSD will capture that. Given a stochastic process $X(t)$, the MSD is defined as

$$MSD(\tau) = \mathbb{E}[(X(\tau) - X(0))^2] \quad (1)$$

(see e.g., [2]). We introduce three common ways to compute the MSD from data. The first two

can be applied to any function of time whereas the last method is applied to stochastic processes. The first two use discretized time data (whether from discrete or continuous data), which is true for our application to cell motion and is typically the case when tracking particle motion.

For the first two methods to approximate the MSD we will assume we have particle positions at evenly spaced time intervals. That is, $x_i = x(t_i)$ for $i = 0, \dots, N$ where $t_i = t_0 + i\Delta t$.

The first of the three methods for computing the MSD is useful if the experimental data available is a sufficiently long time trajectory for a single particle, then the time averaged approximation to the MSD (TAMSD), at lag time $\tau = k\Delta t$ is commonly defined and calculated as follows [3]:

$$TAMSD_1(\tau) = \frac{1}{N - k + 1} \sum_{i=0}^{N-k} |x_{i+k} - x_i|^2, \quad k = 1, 2, \dots, N - 1, N. \quad (2)$$

The advantage of Eq (2) is that for small values of k , there are many displacements, and the data is well averaged. The disadvantage is that when $k > 1$ there is overlap between the displacements, and successive displacements are usually not independent.

If no overlap is allowed between displacements, then the time averaged approximation to the MSD is computed as

$$TAMSD_2(\tau) = \frac{1}{\lfloor N/k \rfloor} \sum_{i=0}^{\lfloor N/k \rfloor - 1} |x_{(i+1)k} - x_{ik}|^2, \quad k = 1, 2, \dots, N - 1, N \quad (3)$$

where $\lfloor \cdot \rfloor$ denotes the integer part. This allows displacements to be uncorrelated for calculations, but if k is large, it is a poor statistical measure due to fewer sample points. Eqs (2) and (3) are the same when $k = 1$. In the subsequent sections of this paper, we typically assume $k = 1$, and this is the method used when calculating the MSD from simulations (which we refer to as the experimental MSD), since our calculations assume there is independence between successive displacements.

If multiple particles of the same type are being tracked over a short period of time, where each particle is equally weighted, then the ensemble averaged approximation to the MSD (EAMSD) at time τ is defined as:

$$EAMSD(\tau) = \frac{1}{P} \sum_{j=1}^P (|x^j(\tau) - x^j(0)|)^2$$

where P is the number of particles and $x^j(\tau)$ is the location of the j -th particle at time τ , and $x^j(0)$ is the referenced position for the j -th particle. When both types of data are available and the system is ergodic (the time average and ensemble average are equivalent for large time) [4], then a simultaneous time and ensemble average is sometimes used, where a time average MSD is computed for each particle and that is then averaged over all the particles. This is especially helpful when lag times are long and improves the statistics [5]. The EAMSD is applied to a stochastic process where x^j are different realizations of the stochastic process. In the centroid model which we analyze in this paper, time is discrete so our data takes the form $X(\omega)_i = X(\omega, i)$ and Δt has no meaning ($\tau = k$). For example if ω is specified then $X(i) = x^j(i)$. In this setting, the EAMSD is easily recognized as an approximation to the definition of the MSD.

The historical importance of MSD came in the year 1905, when Einstein published his Annus Mirabilis (“extraordinary year”) papers, the second of which contained his research and results on Brownian motion [6]. From his work on the diffusion equation in one dimension he

was able to find a linear, time dependent relationship between the MSD and the diffusion coefficient, D , which is a measure of the rate that a particle can move through a fluid that is in thermal equilibrium. The relationship is given by $MSD(\tau) = 2D\tau$ in one dimension and is extended to $MSD(\tau) = 2dD\tau$ for a d -dimensional system. It was a landmark paper and established the value of statistical mechanics in research. (The continuous time stochastic processes in these applications is approximated by discrete-time processes.) The relationship for MSD was further extended to the viscosity of a purely viscous fluid at thermal equilibrium by research simultaneously developed by both Einstein and Sutherland, although Sutherland's contributions were only recognized recently [7]. The relationship between the diffusion coefficient and the viscosity η of a fluid is given by the Stokes-Einstein-Sutherland relation $D = k_B T / (6\pi\eta R_p)$ where k_B is Boltzmann's constant, T is the absolute temperature, and R_p is the radius of a particle, and the particle experiences Stokes drag [6, 8]. Thus, $MSD(\tau) = 2d\tau k_B T / (6\pi\eta R_p)$. Experimentalists now had a way to calculate the diffusion coefficient. (Its extension to cell motion will be discussed in coming paragraphs.)

Further research has shown that the MSD can be used to determine features of the local rheology of non-Newtonian viscoelastic fluids. The complex shear modulus [5], the dynamic moduli [9], and the creep compliance [10] for these fluids can be found using the MSD. A power law tau dependence between the MSD and tau given by $MSD(\tau) = A\tau^\alpha$ is indicative that a particle is moving by nondiffusive transport when $\alpha \neq 1$. It also describes diffusion through a viscoelastic medium [11, 12]. The exponent, α , is referred to as the MSD scaling exponent, and for physical processes $0 \leq \alpha \leq 2$. When $\alpha < 1$ the process is considered subdiffusive, and for $\alpha > 1$, it is superdiffusive. When the MSD exhibits the relationship $MSD(\tau) = 4D\tau + (V\tau)^2$ with V being velocity, the particle exhibits directed motion with diffusive behavior. These different relationships indicate that the MSD, along with the diffusion coefficient, are helpful in revealing the mode of transport, but not all of the mechanisms driving the transport [13].

For living cells, the Stokes-Einstein-Sutherland relation and other equations derived to explain diffusive processes cannot immediately be applied, since living cells use thermal energy and active transport. Under certain conditions, such as active transport inhibition, they are still relevant and can provide information about transport. The time dependent power law is also a useful tool in understanding motion in living cells. Single and two-particle tracking of particles inside a cell have been done on a large number of cell types to find the MSD and hence the MSD scaling exponent [13]. For living cells, if the scaling exponent is in the subdiffusive range, then it may be indicative of a dense intracellular environment and/or there may be numerous reactions and obstacles inside the cell [14]. If the scaling exponent is in the superdiffusive range, then active transport is present [15]. It was also found when tracking whole cells that there is an inverse relationship between the MSD and the stiffness of a cell [16]. This relationship was seen in cancerous cells when the stiffness of the cell decreased as the cell increased in metastatic potential [17]. In summary, the MSD is not only a useful tool to indicate the transport type and mechanics in living cells [13], but in some cases can give information on specific behaviors. Regardless of the application, the MSD is an important statistical measure on a stochastic process.

In this paper we will first discuss the MSD for a simple random walk. We then discuss calculating the MSD for a mathematical model for cell motion. A good estimate for the MSD is found as well as an upper and lower bound for the MSD for this model. We then compare and contrast numerical results found for the simple random walk and our model. Given these methods for estimating the MSD of cell motion with this model, the historical uses of the MSD can be applied to further examine and understand the mechanisms of motion.

Random walks

Some background in simple random walks is needed in order to better understand the model describing cell motion in the next section, so a brief discussion of random walks and the MSD for random walks is given here. A random walk or drunkard's walk was first referred to in 1905 in the journal *Nature* in a discussion between Pearson and Rayleigh, demonstrating the theorem, "the most likely place to find a drunken walker is somewhere near his starting point [18]." Since that time, random walk theory has been studied extensively, impacting many important fields, such as random processes, random noise, stochastic equations and spectral analysis. For a more thorough discussion of random walks in biology, see "Random Walk Models in Biology", by Codling, et.al. [19].

A simple random walk refers to a stochastic process that is the equivalent of a succession of random steps in some finite space or grid. One feature of a random walk is that the jumps are independent. A simple random walk is both time homogeneous ($\mathbb{P}(\mathbf{X}(t) = j \mid \mathbf{X}(0) = a) = \mathbb{P}(\mathbf{X}(s+t) = j \mid \mathbf{X}(s) = a)$) and space homogeneous ($\mathbb{P}(\mathbf{X}(t) = j \mid \mathbf{X}(0) = a) = \mathbb{P}(\mathbf{X}(t) = j + b \mid \mathbf{X}(0) = a + b)$) [20]. Since the process is space homogeneous, we can assume that $\mathbf{X}(0) = \mathbf{0}$, for our purposes. These properties of simple random walks then give that $\mathbb{E}[\|\mathbf{X}(t + \tau) - \mathbf{X}(t)\|^2] = \mathbb{E}[\|\mathbf{X}(\tau) - \mathbf{X}(0)\|^2] = \mathbb{E}[\|\mathbf{X}(\tau)\|^2]$. Since $\mathbf{Var}(\|\mathbf{X}\|) = \mathbb{E}[\|\mathbf{X}\|^2] - \|\mathbb{E}[\mathbf{X}]\|^2$, then $\mathbb{E}[\|\mathbf{X}(\tau) - \mathbf{X}(0)\|^2] = \mathbf{Var}[\|\mathbf{X}(\tau)\|] + \|\mathbb{E}[\mathbf{X}(\tau)]\|^2$. Each $\mathbf{X}(t)$ is the sum of random, independent, identically distributed variables (iids), so $\mathbf{Var}[\|\mathbf{X}(\tau)\|] = \tau \mathbf{Var}[\|\mathbf{X}(1)\|]$ and $\mathbb{E}[\mathbf{X}(\tau)] = \tau \mathbb{E}[\mathbf{X}(1)]$.

Then using the definition of the MSD as defined in the first section,

$$\begin{aligned} MSD(\tau) &= \mathbb{E}[\|\mathbf{X}(\tau) - \mathbf{X}(0)\|^2] \\ &= \tau \cdot \mathbf{Var}[\|\mathbf{X}(1)\|] + \tau^2 \cdot \|\mathbb{E}[\mathbf{X}(1)]\|^2. \end{aligned} \quad (4)$$

This relation shows that the MSD for a simple random walk is a quadratic function in τ . In addition to simple random walks, Eq (4) holds true for any process that is both time and space invariant with \mathbf{X} being the sum of iids.

Eq (4) is the definition for the MSD used in the next section for our calculations to find the MSD for the model of cell motion. Given the expectation and variance, Eq (4) gives us a way to compute the MSD as a function of τ . In a later section, we compare Eq (4) with results from simulations of our model of cell motion (which is not a sum of iids) to see if it has a similar behaviour.

Finding an estimate for the MSD for a model of cell motion

In a paper by John Dallon, et.al. [21], they introduce a mathematical model of individual cell migration. The model specifies discrete focal adhesion (FA) attachment sites with random switching terms for each site. (A FA is a complex that allows a cell to adhere to the extracellular matrix and is integral to amoeboid cell motion.) The random switching terms determine if a FA is attached or detached. The time a FA remains attached or detached is taken from a given probability distribution. A detached site is reattached at an outreach distance from the present cell center. The outreach is taken from a given probability distribution that specifies the distance and location from the centroid where the FA attaches. Forces exerted on the center of the cell by the different FAs are determined by Hooke's Law. Using Newton's second law of motion, and ignoring the acceleration due to the low Reynolds number, all of these forces together with the drag force, which involves velocity, are summed to produce a differential equation model that has the feature of different FAs attaching and detaching randomly and tracks the movement of the cell over time. See Fig 1.

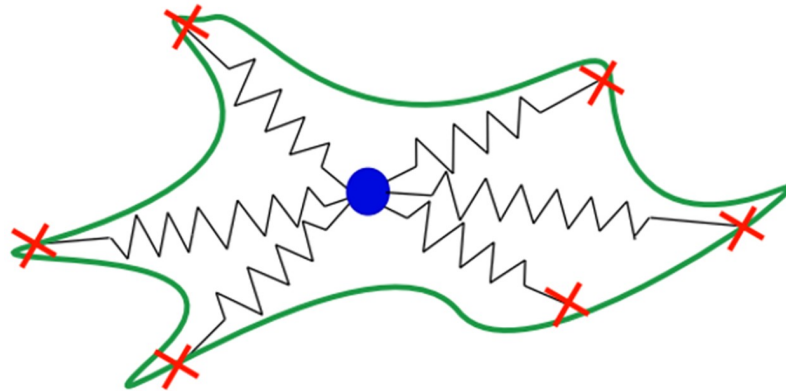


Fig 1. Cell model. This figure from Dallon, et.al. [21] depicts the way the cell is being modeled mathematically. The cell is a center location (nucleus) with attached springs. The other ends of the springs correspond to the different FAs that are attached to the extracellular matrix at “x”.

<https://doi.org/10.1371/journal.pone.0261021.g001>

The equation of motion for the cell location is given by

$$\mu \mathbf{x}' = - \sum_{i=1}^n \alpha_i (\|\mathbf{x} - \mathbf{v}_i\| - \ell_i) \frac{\mathbf{x} - \mathbf{v}_i}{\|\mathbf{x} - \mathbf{v}_i\|} \psi_i(t),$$

where \mathbf{x} is the location of the center of the cell, α_i is the constant for the i th spring with resting length ℓ_i , \mathbf{v}_i is the location of the i th FA, n is the total number of FAs, and $\psi_i(t)$ is the state of the i th FA at time t . If $\psi_i(t) = 0$, then the i th FA is detached at time t , and if $\psi_i(t) = 1$, then the i th FA is attached at time t .

The equation for the location of the i th FA \mathbf{v}_i when it is going from detached to attached is given by

$$\mathbf{v}_i(t) = \lim_{y \nearrow a_{p,i}} \mathbf{x}(y) + \boldsymbol{\eta}^{p,i} \text{ for } a_{p,i} \leq t < a_{p+1,i}.$$

For each i the sequence $\{a_{p,i}\}$ of random variables are the times when ψ_i makes the transition from 0 to 1. The vectors $\boldsymbol{\eta}^{p,i}$ are independent, identically distributed random vectors with respect to a distribution ν on \mathbb{R}^2 . (The vector $\boldsymbol{\eta}^{p,i}$ is the outreach. The two superscripts denote the time sequence and FA respectively.) Although the equations of motion are independent of the location of the FA when it is detached, for convenience we assume the location does not change until it reattaches.

In a further paper [22], the differential equation model is approximated heuristically by a problem that tracks the centroid of the cell, \mathbf{c}^j . This new problem is motivated by informally considering the limit of the differential equation model as the cell spring constants become very large. In this limit, the cell nucleus jumps from centroid to centroid. This model is a discrete-time jump process and is the model for which we estimate the MSD.

It is described in the following manner. Let j denote the number of binding events (attach or detach events) that have occurred and n the number of FAs. The equation describing \mathbf{c}^j is

$$0 = \sum_{i=1}^n \alpha_i (\mathbf{c}^j - \mathbf{v}_i^j) \psi_i^j$$

where \mathbf{v}_i^j is the location of the i th attachment site at stage j , α_i is the spring constant for the i th attachment and ψ_i^j is either 1 if the i th attachment site is attached at event j or 0 if the i th

attachment site is detached at event j . If the j th event is the attachment of \mathbf{v}_i , then its location is governed by

$$\mathbf{v}_i^j = \mathbf{c}^{j-1} + \boldsymbol{\eta}^j.$$

The vectors $\boldsymbol{\eta}^j$ are independent, identically distributed random vectors with respect to a distribution ν on \mathbb{R}^2 . (For our simulations, the outreach is specified by a length and an angle.)

Each attached FA has a certain probability p of changing status, and each detached FA has a probability rp of changing status, where $r > 0$. If k is the number of attached FAs, then $kp + (n - k)rp = 1$. Thus for given k , the probability of going from k to $k + 1$ attachments is given by rp_k with

$$p_k = \frac{1}{k + (n - k)r} \tag{5}$$

as found in [22]. The probability p_k is the probability of going from k to $k - 1$ attachments.

The behavior of the system can be quite complex unless we restrict the initial conditions to be compatible with a steady state distribution of the number of attached FAs. It was previously shown that this distribution is a globally attracting steady state, so if the process is run long enough this is not very restrictive.

More precisely, we restrict the initial conditions to come from a distribution ρ , which is a distribution on the Borel sets of the possible cell states, $\mathcal{B}(X)$, where

$$X := \{((\psi_1, \dots, \psi_n), (\mathbf{v}_1, \dots, \mathbf{v}_n), \mathbf{c}) \in \{0, 1\}^n \times (\mathbb{R}^2)^n \times \mathbb{R}^2 : \sum_{i=1}^n \psi_i(\mathbf{v}_i - \mathbf{c}) = 0\}.$$

(We give X the product topology with the discrete topology on $\{0, 1\}$ and the standard topology on \mathbb{R}^2 .) We put a further restriction on ρ , such that the probabilities of a projection of X onto the number of attachments $|\psi|$ associated with any given configuration is consistent with the steady state distribution. This is given by the equation

$$\rho((\psi_1, \dots, \psi_n) \times (\mathbb{R}^2)^n \times \mathbb{R}^2) \cap X = \pi_{|\psi|}$$

for every $(\psi_1, \dots, \psi_n) \in \{0, 1\}^n$ with $\pi_{|\psi|}$ being the probability of the projected steady state. This steady state was computed in Dallan, et al. [22] and is given by

$$\pi_k = \frac{r^{k-1} \binom{n-1}{k-1}}{2(1+r)^{n-1} k} [k + (n-k)r], \tag{6}$$

the probability of being in the projected state of k attachments for any configuration with $0 < k \leq n$ and $\pi_0 = 1/(2(1+r)^{n-1})$.

Analysis of the centroid model by the authors in [22] produced an explicit formula for the first moment, given by

$$\mathbb{E}_\rho[\mathbf{c}^{j+1} - \mathbf{c}^j] = \left(1 + \sum_{k=1}^n \frac{\left(r^{k-1}(1-r) \binom{n-1}{k-1} + r^k \binom{n}{k} \right) r(n-k)}{(k+r(n-k))(k+1)} \right) \frac{\mathbb{E}_\nu[\boldsymbol{\eta}]}{2(1+r)^{n-1}} \tag{7}$$

where ρ is a probability measure on the Borel sets of the state space as defined above [22].

Table 1. Centroid model (n = 2) for j ≥ 1. This table shows the probabilities of being in certain states, the probabilities of changing to different states with the number of possibilities for those changes, and the centroid location change.

| Probability of State | $ \psi^j $ | $ \psi^{j+1} $ | Possibilities | $\mathbf{c}^{j+1} - \mathbf{c}^j$ | Probability of Attach/Detach |
|----------------------------|------------|----------------|----------------|-------------------------------------|------------------------------|
| $\pi_0 = \frac{1}{2(1+r)}$ | 0 | 1 | $\binom{2}{1}$ | $\boldsymbol{\eta}^{j+1}$ | $rp_0 = \frac{1}{2}$ |
| $\pi_1 = \frac{1}{2}$ | 1 | 2 | $\binom{1}{1}$ | $\frac{\boldsymbol{\eta}^{j+1}}{2}$ | $rp_1 = \frac{r}{1+r}$ |
| | 1 | 0 | $\binom{1}{1}$ | 0 | $p_1 = \frac{1}{1+r}$ |
| $\pi_2 = \frac{r}{2(1+r)}$ | 2 | 1 | $\binom{2}{1}$ | $\pm \frac{\boldsymbol{\eta}^j}{2}$ | $p_2 = \frac{1}{2}$ |

<https://doi.org/10.1371/journal.pone.0261021.t001>

It is noted that the MSD of the centroid in this setting changes the meaning of τ from a time shift to an event shift. We work to determine a similar formula for the MSD of one event shift ($\tau = 1$), i.e. $\mathbb{E}_\rho[\|\mathbf{c}^{j+1} - \mathbf{c}^j\|^2]$ for any $j \geq 0$.

Case: n = 1

Consider $n = 1$ with $|\psi^j| = 0$ where $|\psi^j|$ is the number of attached sites at time j , and compute $\mathbf{c}^{j+1} - \mathbf{c}^j$. Since the process is space invariant, then we can assume that $\mathbf{c}^j = 0$. In this case, the difference $\mathbf{c}^{j+1} - \mathbf{c}^j$ would be the outreach from the centroid on the next step, $\boldsymbol{\eta}^{j+1}$. For $|\psi^j| = 1$, the only possibility for the next event would be going from one attachment to no attachments. (We assume that if all the FAs detach, then the location of the centroid does not move.) In this case, the centroid does not move, so $\mathbf{c}^{j+1} - \mathbf{c}^j = 0$. Those two cases then give the only possible values for the random variables, $\mathbf{c}^{j+1} - \mathbf{c}^j$, in the stochastic process when $n = 1$. Thus for $n = 1$, $\mathbb{E}_\rho[\|\mathbf{c}^{j+1} - \mathbf{c}^j\|^2] = \pi_0 rp_0 \mathbb{E}_v[\|\boldsymbol{\eta}\|^2] + \pi_1 p_1 \mathbb{E}_v[0] = \frac{1}{2} r \left(\frac{1}{r}\right) \mathbb{E}_v[\|\boldsymbol{\eta}\|^2] = \frac{\mathbb{E}_v[\|\boldsymbol{\eta}\|^2]}{2}$

Case: n = 2

For $n = 2$, $\mathbf{c}^{j+1} - \mathbf{c}^j$ (for any $j \geq 1$) can be computed for all scenarios of FA attachments/detachments. See Table 1. A visualization for $n = 2$ can be found in Fig 2. Note that the open dots indicate a detached adhesion site and a black dot represents an attached adhesion site. An “x” indicates the centroid.

For $n = 2$, the MSD can be computed by substituting the values of the random variables and associated probabilities found in [22] into Eq (1) with $\tau = 1$ and is $\frac{\mathbb{E}_v[\|\boldsymbol{\eta}\|^2]}{2(1+r)} (1 + \frac{r}{2})$.

Case: n > 2

For $n > 2$, the initial condition is a cell that has not yet formed attachments to the substrate, i.e., initially no FAs are attached. In order to find a good estimate for the theoretical MSD, we considered two features of the model: (1) Only one event happens at a time. and (2) The probability that a single FA (focal adhesion) remains attached for a long period of time is small.

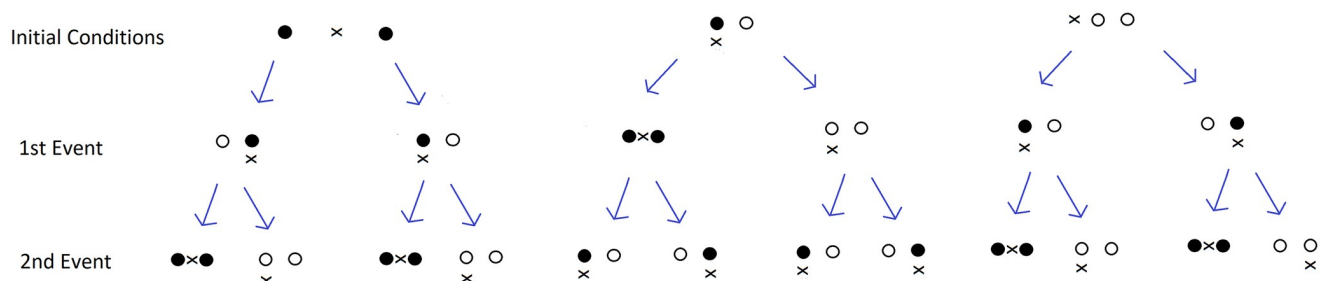


Fig 2. Visualization of the centroid model (n = 2). The left column shows the three possible initial conditions: No attached FAs, one attached FA and 2 attached FAs (in any configuration). The arrows point to possible transitions. Distance is measured vertically. The open dots indicate a detached adhesion site and a black dot represents an attached adhesion site. An “x” indicates the centroid. When the centroid is in the same position as a dot, then it is indicated to the right of the dot.

<https://doi.org/10.1371/journal.pone.0261021.g002>

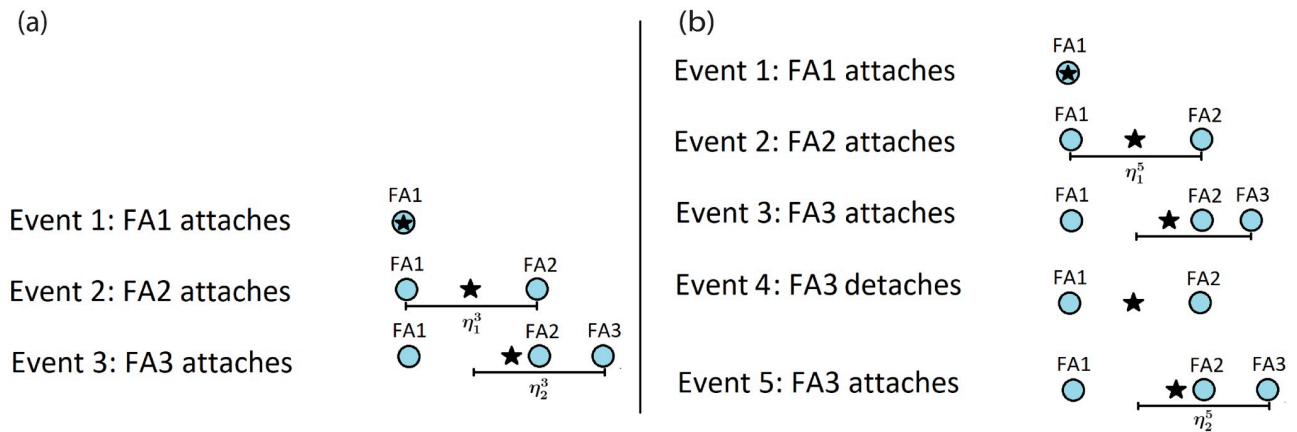


Fig 3. Visualization of sequential attachment configuration with two different histories. For simplicity, the visualization is in one dimension. In panel (a), the first FA (FA1) attaches, denoted by the blue circle. The location of the centroid c^1 is denoted by the star. The next FA (FA2) attaches at distance η_1 from the centroid. The new centroid is calculated, c^2 . The third FA (FA3) attaches at a distance η_2 from the centroid. A new centroid is calculated, c^3 . Panel (b) shows a different history that still has the same sequential configuration as panel (a) because they share a common sequential creation story. The sequential creation story for event 5 has the sequential attachments happening at Event 2 and Event 5.

<https://doi.org/10.1371/journal.pone.0261021.g003>

These two features imply that probabilistically, the FAs will be close together throughout the process. To find a computable estimate for the MSD, we introduce the concept of sequential attachments. By sequential attachments, we mean that for any $k \leq n$, the k attachments are sequential if they are in a configuration that can be arrived at by starting with a centroid and no attachments and then attaching one FA at a random outreach (ν -distributed) from the centroid. Then the new centroid location is computed and another FA attaches at a random outreach from the centroid. Each new FA attaches in this same way until k are attached. The FAs are also considered sequential if they are in the configuration described above whether or not they arrived in that manner. In other words, FAs are in a sequential configuration if they have a sequential creation story. See Fig 3. (Not all attachments are in a sequential configuration. For example, in panel (b) of Fig 3, if at Event 4, FA2 detaches, then the configuration is no longer sequential.) Assuming sequential attachments assures that the FAs are close together and makes it possible to compute the displacement of the centroid when a detach event occurs.

The values for $c^{j+1} - c^j$ are computed for $n = 5$ as shown in Table 2. The values for an attach event are valid for any configuration in the state space, but the ones for a detach event are only valid if the configuration is a sequential attachment. For the purposes of finding an estimate for the MSD, we assign the full probabilities of the state space to both attach and detach events, even though the random variable for the detach events is only for a sequential configuration. Since the FAs in general are close together, and a sequential configuration achieves this proximity, then assigning the full probability to this case is reasonable.

In general, for n total FAs

$$c^{j+1} - c^j = \frac{\eta_k^{j+1}}{k} \tag{8}$$

when going from $|\psi^j| = k - 1$ to $|\psi^{j+1}| = k$ attached sites (an attach event) with $1 \leq k \leq n$, where the superscripts are an event counter, and the subscript on η for an attach event is the outreach to the k th FA. When going from $|\psi^j| = k$ to $|\psi^{j+1}| = k - 1$ attached sites (a detach event) with $2 \leq k \leq n$ then the subscript on η is the outreach order in the creation story of the sequential configuration. In this case, there are $\binom{k}{1}$ possibilities for $c^{j+1} - c^j$. For $\ell = 0$ to $k - 2$, the ℓ th

Table 2. Centroid model (n = 5).

| Probability of State | $ \psi^j $ | $ \psi^{j+1} $ | Possibilities | $c^{j+1} - c^j$ | Probability of Attach/Detach |
|---------------------------------------|------------|----------------|----------------|---|------------------------------|
| $\pi_0 = \frac{1}{2(1+r)^4}$ | 0 | 1 | $\binom{5}{1}$ | η_1^{j+1} | $rp_0 = \frac{1}{5}$ |
| $\pi_1 = \frac{1+4r}{2(1+r)^4}$ | 1 | 2 | $\binom{4}{1}$ | $\frac{\eta_2^{j+1}}{2}$ | $rp_1 = \frac{r}{1+4r}$ |
| | 1 | 0 | $\binom{4}{1}$ | 0 | $p_1 = \frac{1}{1+4r}$ |
| $\pi_2 = \frac{2r(3r+2)}{2(1+r)^4}$ | 2 | 3 | $\binom{3}{1}$ | $\frac{\eta_3^{j+1}}{3}$ | $rp_2 = \frac{r}{2+3r}$ |
| | 2 | 1 | $\binom{2}{1}$ | $\pm \frac{\eta_2^j}{2}$ | $p_2 = \frac{1}{2+3r}$ |
| $\pi_3 = \frac{2r^2(2r+3)}{2(1+r)^4}$ | 3 | 4 | $\binom{2}{1}$ | $\frac{\eta_4^{j+1}}{4}$ | $rp_3 = \frac{r}{3+2r}$ |
| | 3 | 2 | $\binom{3}{1}$ | $\frac{1}{2} \left(\frac{\eta_2^j}{2} + \frac{\eta_3^j}{3} \right)^*$ | $p_3 = \frac{1}{3+2r}$ |
| | | | | $\frac{1}{2} \left(-\frac{\eta_2^j}{2} + \frac{\eta_3^j}{3} \right)^*$ | |
| | | | | $-\frac{\eta_3^j}{3}$ * | |
| $\pi_4 = \frac{r^3(r+4)}{2(1+r)^4}$ | 4 | 5 | $\binom{1}{1}$ | $\frac{\eta_5^{j+1}}{5}$ | $rp_4 = \frac{r}{4+r}$ |
| | 4 | 3 | $\binom{4}{1}$ | $\frac{1}{3} \left(\frac{\eta_2^j}{2} + \frac{\eta_3^j}{3} + \frac{\eta_4^j}{4} \right)^*$ | $p_4 = \frac{1}{4+r}$ |
| | | | | $\frac{1}{3} \left(-\frac{\eta_2^j}{2} + \frac{\eta_3^j}{3} + \frac{\eta_4^j}{4} \right)^*$ | |
| | | | | $\frac{1}{3} \left(-\frac{2\eta_3^j}{3} + \frac{\eta_4^j}{4} \right)^*$ | |
| | | | | $-\frac{\eta_4^j}{4}$ * | |
| $\pi_5 = \frac{r^4}{2(1+r)^4}$ | 5 | 4 | $\binom{5}{1}$ | $\frac{1}{4} \left(\frac{\eta_2^j}{2} + \frac{\eta_3^j}{3} + \frac{\eta_4^j}{4} + \frac{\eta_5^j}{5} \right)^*$ | $p_5 = \frac{1}{5}$ |
| | | | | $\frac{1}{4} \left(-\frac{\eta_2^j}{2} + \frac{\eta_3^j}{3} + \frac{\eta_4^j}{4} + \frac{\eta_5^j}{5} \right)^*$ | |
| | | | | $\frac{1}{4} \left(-\frac{2\eta_3^j}{3} + \frac{\eta_4^j}{4} + \frac{\eta_5^j}{5} \right)^*$ | |
| | | | | $\frac{1}{4} \left(-\frac{3\eta_4^j}{4} + \frac{\eta_5^j}{5} \right)^*$ | |
| | | | | $-\frac{\eta_5^j}{5}$ * | |

The superscripts are an event counter, and the subscript on η for an attach event is the outreach to the k th FA and for a detach event is the outreach order in the creation story of the sequential configuration.

*The starred values are only valid for sequential attachments. (For the purposes of finding an estimate for the MSD, the probabilities on the table are for the entire state space even though the random variables for a detach event are only valid for a sequential configuration.)

<https://doi.org/10.1371/journal.pone.0261021.t002>

possibility is

$$\frac{1}{k-1} \sum_{i=0}^{k-(\ell+2)} \frac{\eta_{k-i}^j}{k-i} - \frac{\ell \eta_{\ell+1}^j}{(\ell+1)}. \tag{9}$$

The last possibility is

$$-\frac{\eta_k^j}{k}. \tag{10}$$

Each of these possibilities, corresponds to a particular site in the creation story detaching on event $j + 1$.

If we consider each configuration of $\mathbf{c}^{j+1} - \mathbf{c}^j$, where the number of attachments is known as is the nature of the next event (attach or detach), and consider the possible values of that difference as random variables which depend only on the distribution ν , then we can determine expectations. By using Eqs (8), (9) and (10), we can determine expectations with respect to ν that contribute to an MSD estimate of the full state space.

Configuration attach:

We find $\mathbb{E}_\nu[\|\mathbf{c}^{j+1} - \mathbf{c}^j\|^2]$ for any number of attachments, k , with the next event being an attachment by using Eq (8). Thus, for $|\psi^j| = k - 1$ and $|\psi^{j+1}| = k$ and $1 \leq k \leq n - 1$ then

$$\mathbb{E}_\nu[\|\mathbf{c}^{j+1} - \mathbf{c}^j\|^2] = \frac{\mathbb{E}_\nu[\|\boldsymbol{\eta}^{j+1}\|^2]}{k^2} = \frac{\mathbb{E}_\nu[\|\boldsymbol{\eta}\|^2]}{k^2} \tag{11}$$

where the norm is defined in terms of the inner product.

Configuration detach (assuming sequential configuration):

Similarly, we find $\mathbb{E}_\nu[\|\mathbf{c}^{j+1} - \mathbf{c}^j\|^2]$ for any number of attachments k with the next event being a detachment using Eqs (9) and (10). Thus, for $|\psi^j| = k$ and $|\psi^{j+1}| = k - 1$, with $2 \leq k \leq n$ ($\mathbf{c}^{j+1} - \mathbf{c}^j = 0$ when $k = 1$), then

$$\mathbb{E}_\nu[\|\mathbf{c}^{j+1} - \mathbf{c}^j\|^2] = \frac{1}{(k - 1)^2} \sum_{i=1}^{k-1} \frac{i \mathbb{E}_\nu[\|\boldsymbol{\eta}\|^2]}{i + 1}. \tag{12}$$

Using the expectations found in Eqs (11) and (12), we derive an estimate for the MSD. For $n > 1$ adhesion sites the estimated theoretical MSD with respect to the initial distribution, ρ , that is compatible with the projected steady state found in [22], assuming only a sequential configuration for a detach event with a full state space probability for all events, is given by

$$MSD(1) = \mathbb{E}_\rho[\|\mathbf{c}^{j+1} - \mathbf{c}^j\|^2] \approx \frac{\mathbb{E}_\nu[\|\boldsymbol{\eta}\|^2]}{2(1+r)^{n-1}} \left(1 + \sum_{k=1}^{n-1} \binom{n-1}{k} \frac{r^k}{(k+1)^2} + \binom{n-1}{k} \frac{r^k}{(k+1)(k^2)} \sum_{i=1}^k \frac{i}{i+1} \right). \tag{13}$$

To find this estimate for the MSD, which we will refer to as AMSD (approximate MSD), we first find the contribution to the expectation due to the attachment events. Eq (11) is multiplied by π_k and by rp_k (the probability of going from k to $k + 1$ attachments) and by the number of possibilities $n - k$, for $1 \leq k \leq n - 1$, given by $\pi_k rp_k (n - k) \frac{\mathbb{E}_\nu[\|\boldsymbol{\eta}\|^2]}{k^2}$. Summing these products over k gives the first summation, or second term, in Eq (13) with $\frac{\mathbb{E}_\nu[\|\boldsymbol{\eta}\|^2]}{2(1+r)^{n-1}}$ being factored out from all terms. The first term (“1”) is when $k = 0$. In order to find an estimate for the detach events, we multiply Eq (12), the equation for sequential attachments, by π_k and p_k (the full probability of going from k to $k - 1$ attachments) with $1 \leq k \leq n$, given by $\pi_k p_k \frac{1}{(k-1)^2} \sum_{i=1}^{k-1} \frac{i \mathbb{E}_\nu[\|\boldsymbol{\eta}\|^2]}{i+1}$. Summing over all k , with an appropriate change of indices yields the third term in Eq (13). In summary, the first term is for the attachment event when $k = 0$, the second term is the sum over all other attachment events and the third term is the sum over detachment events.

Some numerical simulations were conducted to see how closely this formula compares to the experimental MSD (for the full state space—not just sequential attachments), where we assume that $\tau = 1$. For 10,000,000 iterations and fixed r , the MSD was computed and compared to the number of FAs.

The numerical simulation to determine the experimental MSD begins with the location of the FAs in a circle equally spaced around the origin at a random distance from 0 to 10 with all FAs attached. It proceeds as follows:

1. Generate a number from the standard uniform distribution.
2. If this number is less than $r * p * (\text{number of detached FAs})$ where $p = 1/(|\psi| + (n - |\psi|)r)$, then the event is an attachment. Using MATLAB's random number generator, a random detached FA is selected and its length and angle of outreach is chosen from a random distribution, and it is attached at the chosen length and angle from the present centroid.
3. If #2 is not true, then the event is a detachment. A random attached FA is selected and detached.
4. The new location of the centroid is computed.
5. The location of the centroid is not recorded until a preset amount of events have happened, the burn-in period. (This is done to "wash out" the initial conditions.)
6. The simulation continues until the specified number of events has happened.
7. The data file of the centroid locations at each event is then used to compute the $MSD(1)$.

The graph of the AMSD from Eq (13) was also computed for fixed r and number of FAs and was juxtaposed on the same graph, (see Fig 4). As seen from the graph, Eq (13), is a good estimate for the MSD.

Lower bound

We now find a lower bound for the MSD. The results will help us understand why the AMSD is such a good approximation. We discuss the insights gained in the section entitled "Discussion on the MSD Estimate".

In order to find a lower bound for the MSD, given n total FAs, we used the random variable values found in Eqs (8), (9) and (10), and for values that are unknown we used 0. For

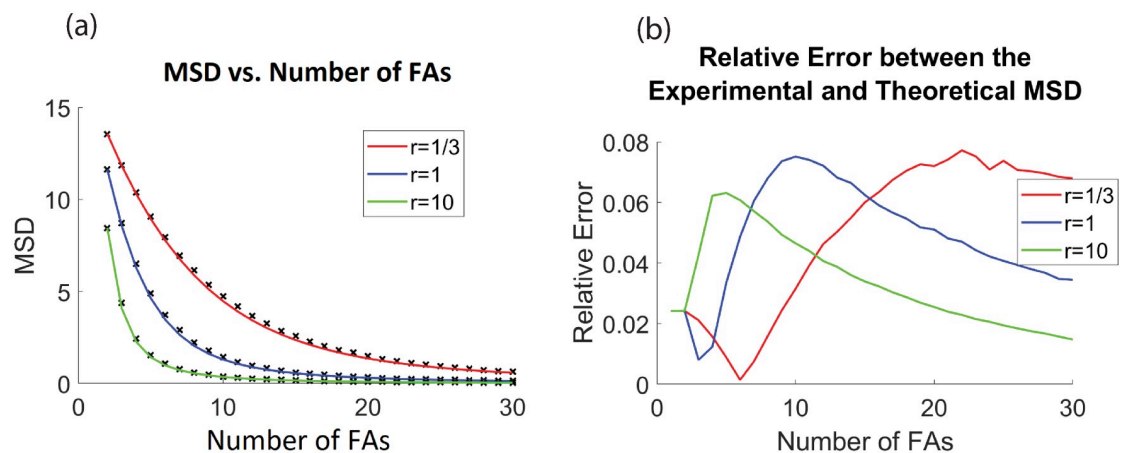


Fig 4. Comparison between the experimental MSD and the estimated theoretical MSD. For the left panel, the experimental MSD ($TAMSD_2$) is computed from a simulated trajectory of 10,000,000 points and is marked with a black "x" for different values of FAs. It is compared to the AMSD found in Eq (13), given by the solid lines. The right panel shows the relative error between $TAMSD_2$ and AMSD for different values of r . For this simulation and all reported simulations the angle of η (the outreach vector) is from -30 to 30 degrees, and the length of η is from 0 to 10. The centroid model is a space and time homogeneous process, thus $TAMSD_2 \approx MSD$.

<https://doi.org/10.1371/journal.pone.0261021.g004>

random variable values from Eqs (9) and (10) (a detach event) we used the probability of being in a sequential configuration when the creation story and the actual history coincide. The probability of starting with no attachments is π_0 . The probability of attaching one FA is rp_0 multiplied by the number of possibilities of FAs to attach, which is n . The probability of attaching another FA is rp_1 multiplied by the number of possibilities, $n - 1$. We continue until we attach the k th FA, which has probability $rp_{k-1}(k - 1)$. Multiplying all of these probabilities together and then multiplying by $p_k(k)$ (the probability of being in the state of k attachments and detaching one of them) gives the probability of being in this particular sequential configuration of k attachments and then detaching one of the FAs. Thus, given n FAs, the probability of being in this particular sequential state of k attachments and then detaching one of them is

$$\begin{aligned}
 P_k^d(r) &= \pi_0(rp_0n)(rp_1(n - 1)) \dots (rp_{k-1}(n - (k - 1)))(p_kk) \\
 &= \pi_0kr^k p_0 p_1 \dots p_k \left(\frac{n!}{(n - k)!}\right) \\
 &= \frac{1}{2(1 + r)^{n-1}} \left(\frac{r}{nr}\right) \left(\frac{r}{1 + (n - 1)r}\right) \dots \\
 &\quad \left(\frac{r}{(k - 1) + (n - (k - 1))r}\right) \left(\frac{k}{k + (n - k)r}\right) \left(\frac{n!}{(n - k)!}\right)
 \end{aligned} \tag{14}$$

where $1 \leq k \leq n$. Using these adjusted probabilities for the detach event random variables, we can obtain a lower bound (LB) for the MSD, and it is given by

$$\begin{aligned}
 LB &= \frac{\mathbb{E}_v[\|\boldsymbol{\eta}\|^2]}{2(1 + r)^{n-1}} \left(1 + \sum_{k=1}^{n-1} \left(\binom{n-1}{k} \frac{r^k}{(k+1)^2} + \right. \right. \\
 &\quad \left. \left. r^k(k+1) \left(\prod_{i=1}^{k+1} \frac{1}{i + (n-i)r}\right) \binom{n}{k} k! \frac{1}{k^2} \sum_{j=1}^k \frac{j}{j+1}\right)\right).
 \end{aligned} \tag{15}$$

The first two terms are the same as in Eq (13). The third term is found by using the expectations for a detachment event (assuming sequential configuration) computed in Eq (14), but using the probabilities from Eq (14). A graph of how it compares to the experimental MSD and estimated theoretical MSD can be seen in Fig 5.

Upper bound

In this section, we seek to find an upper bound for the MSD. In order to find a tight upper bound, it would require an extensive partitioning of the state space, and the possibilities are too numerous to be practicable. Instead, we find a worst-case scenario for the centroid displacement and show that it can be attained, but is highly improbable. Although our upper bound is not ideal, the methods to obtain it are instructive, and at least the existence of an upper bound is proved.

To postulate on the maximum value for $\|\mathbf{c}^{j+1} - \mathbf{c}^j\|$, when event $j + 1$ is a detachment, we describe a “worst-case scenario” event. We start with our initial condition assumption of the centroid at the origin with no attachments. Assume the first FA, v_1 , attaches at the origin. For simplicity and to obtain a maximum combined outreach, we assume all incremental outreaches occur in one dimension in the positive direction. The next FA, v_2 , attaches at a maximum outreach, η_{max} from the origin. Let each subsequent outreach be at a maximum outreach from the previously attached FA until all n FAs are attached, the location of the i th FA given by v_i , with $v_1 = 0$ and $v_n = \eta_{max}(n - 1)$. (This is a maximum outreach scenario that

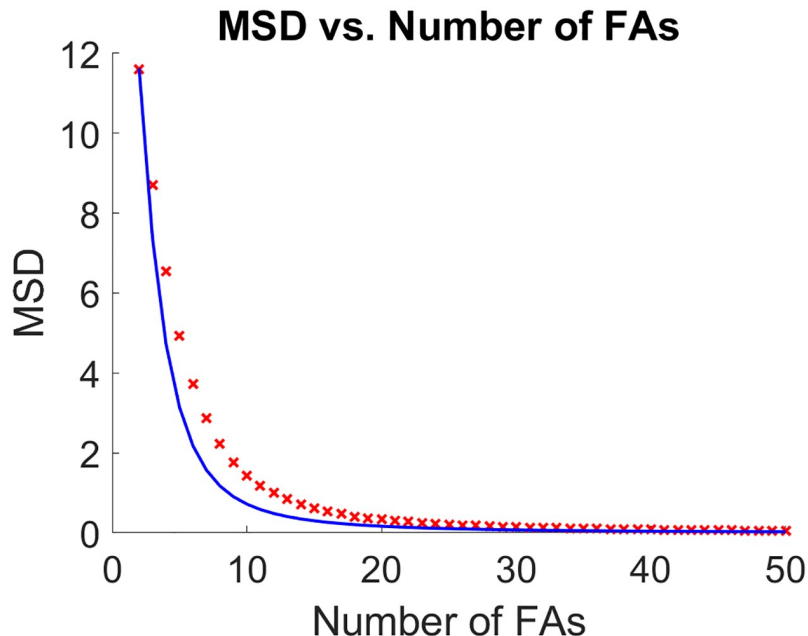


Fig 5. Lower bound for the MSD. Experimental MSD (red x's) compared against the lower bound for the MSD (blue line) found in Eq (15). For this graph, a trajectory of 10,000 data points was used for the experimental MSD with $r = 1$. <https://doi.org/10.1371/journal.pone.0261021.g005>

is more than the actual model, since the outreach in the model for each new attaching FA is from the centroid.) By fixing v_1 at 0 for all events up through j , and allowing v_d to detach ($v_d \in \{v_i | 1 \leq i \leq n\}$) for the event $j + 1$ ($j > n$) we can find an upper bound for any $\|c^{j+1} - c^j\|$.

$$\begin{aligned} \|c^{j+1} - c^j\| &= \frac{\sum_{i=1}^n v_i - v_d}{n-1} - \frac{\sum_{i=1}^n v_i}{n} = \frac{\sum_{i=2}^n v_i - nv_d}{n(n-1)} \\ &\leq \frac{\eta_{max}(n-1)(n-1)}{n(n-1)} = \frac{\eta_{max}(n-1)}{n}. \end{aligned} \tag{16}$$

where the values after the inequality come from taking the max value for all v_i and taking the minimum value of 0 for v_d .

In general, for k attachments ($1 \leq k \leq n$) we find an upper bound for the displacement by using the upper bound configuration found in Eq (16), i.e. all nonzero FAs are $\eta_{max}(n-1)$ units away from the origin. So for $j > n$

$$\|c^{j+1} - c^j\| \leq \frac{\eta_{max}(n-1)(k-1)}{k-1} - \frac{\eta_{max}(n-1)(k-1)}{k} = \frac{\eta_{max}(n-1)}{k}. \tag{17}$$

We now show analytically that the maximum displacement bound found in Eq (16) can be achieved in the limit. Given n FAs, there is a linear recurrence relation for the location of the next FA, given the location of the previous $n - 1$ FAs, where each x_i is the location of a FA and $t \geq n$.

$$x_t = \frac{(x_{t-1} + x_{t-2} + \dots + x_{t-n+2} + x_1)}{n-1} + \eta_{max}$$

or

$$x_t = \frac{(x_{t-1} + x_{t-2} + \dots + x_{t-n+2})}{n-1} + \eta_{max} \tag{18}$$

since $x_1 = 0$.

The steady state of this equation is found by setting all values of x to x^* and solving for x^* . The steady state is then $x^* = \eta_{max}(n-1)$. In order to find if this is an attracting steady state, let $y_t = x_t - x^*$, and Eq (18) becomes

$$y_t = \frac{(y_{t-1} + y_{t-2} + \dots + y_{t-n+2})}{n-1}. \tag{19}$$

The characteristic equation for this recurrence relation is

$$(n-1)\lambda^{n-2} = \lambda^{n-3} + \dots + \lambda + 1. \tag{20}$$

For ease of computation consider the equivalent system

$$(k+1)\lambda^k = \lambda^{k-1} + \dots + \lambda + 1$$

where $k+1 = n-1$. Thus the characteristic polynomial is $\lambda^k - \frac{\lambda^{k-1}}{k+1} - \dots - \frac{\lambda}{k+1} - \frac{1}{k+1}$.

By Descartes rule of signs, we know that the polynomial has exactly one positive real root. Since one and negative one are not roots of the polynomial, the upper and lower bound theorem for real roots of polynomials says that all of the real roots lie between negative one and one. In particular, the unique positive root, call it ζ , must be between 0 and 1, i.e. $0 < \zeta < 1$. Further analysis shows that $x^k - \frac{x^{k-1}}{k+1} - \dots - \frac{x}{k+1} - \frac{1}{k+1} < 0$ or $x^k \leq \frac{x^{k-1}}{k+1} + \dots + \frac{x}{k+1} + \frac{1}{k+1}$ for all values of $0 \leq x < \zeta$ and $x^k - \frac{x^{k-1}}{k+1} - \dots - \frac{x}{k+1} - \frac{1}{k+1} \geq 0$ for $x \geq \zeta$. Let z_0 be a complex root of the characteristic polynomial, then $z_0^k - \frac{z_0^{k-1}}{k+1} - \dots - \frac{z_0}{k+1} - \frac{1}{k+1} = 0$. Using the triangle inequality, then $|z_0|^k \leq \frac{|z_0|^{k-1}}{k+1} + \dots + \frac{|z_0|}{k+1} + \frac{1}{k+1}$. This implies that $0 < |z_0| < \zeta < 1$. Since z_0 was arbitrary, then all of the complex roots of the characteristic polynomial have modulus less than one. Therefore, all roots of the characteristic polynomial lie within the unit circle in the complex plane, showing that the steady state, $x^* = \eta_{max}(n-1)$ is attracting, and the system will converge to it, since it is the only steady state. As the system approaches the steady state, then the displacement is maximal, and by extending to higher dimensions is given by,

$$\lim_{j \rightarrow \infty} \|c^{j+1} - c^j\| = \frac{\eta_{max}(n-1)(n-1)}{n-1} - \frac{\eta_{max}(n-1)(n-1)}{n} = \frac{\eta_{max}(n-1)}{n} \tag{21}$$

for n total FAs, which is the value seen in Eq (16).

Since the upper bound of the displacement found in Eq (16) can be obtained in the limit (Eq (21)), we now use the results found in Eqs (16) and (17) to find an upper bound for the MSD. We partition the state space into three parts: $\{F_k^a\}$, $\{\tilde{F}_k^d\}$ and $\{F_k^d\}$. Each F_k^a , $0 \leq k \leq n-1$, represents arriving to a state of k attachments from any configuration and then attaching. Each \tilde{F}_k^d , $1 \leq k \leq n$ represents arriving to the state of k attachments from a sequential configuration and then detaching. Each F_k^d , $1 \leq k \leq n$, represents arriving to a state of k attachments from a non-sequential configuration and then detaching. We use the known values and associated probabilities for F_k^a , and we use the RV values in Eqs (9) and (10) with probabilities from (14) for \tilde{F}_k^d in the computation of the MSD upper bound. We use the results from Eq (17), as a RV upper bound for the event of arriving at k attachments from a non-sequential configuration. For the upper bound for the probabilities in this case, we use $k\pi_k p_k - \pi_0 k r^k p_0 p_1 \dots p_k \left(\frac{n!}{(n-k)!}\right)$ (sequential probability from (14) subtracted from the probability of being in a state of k

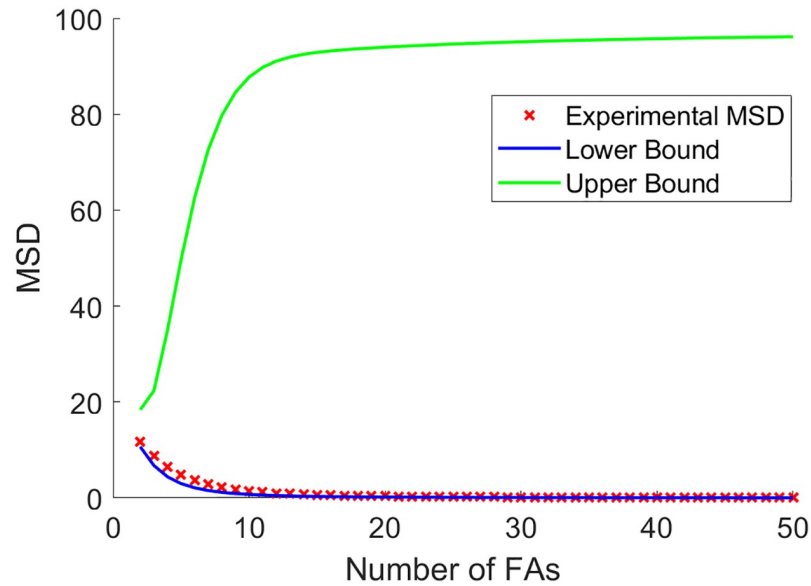


Fig 6. Upper bound for the MSD. An upper bound for the MSD is found using the values for the centroid displacement found in Eq (17). A trajectory of 100,000 data points with $r = 1$ was used to approximate the MSD.

<https://doi.org/10.1371/journal.pone.0261021.g006>

attachments and then detaching). The resultant upper bound can be seen in Fig 6. For $\{F_k^d\}$, since we use a rare event for the upper bound of the displacement (one FA staying attached for a long time), and multiply it by a large probability, then this is the best estimate for an upper bound that can be found without partitioning the space into the many, many ways that the FAs can arrive at a state of k attachments and then have one FA detach.

MSD as a function of τ

Thus far we have considered $MSD(1)$ for the centroid model. We now turn our attention to $MSD(\tau)$ where τ is an integer and represents the number of binding events between the centroid locations being considered. As mentioned earlier, for a random walk, the shape of the MSD curve can reveal if the process is purely diffusive (linear) or has a directed component (quadratic) as can be seen in Eq (4). Eq (4) is valid for more than random walks. It is valid for any process that is both time and space invariant and is the sum of iids. The centroid process we are modeling is both time and space invariant, but as Table 1 indicates, the location of the centroid is not a sum of iids. Since the state space is the location of the centroid and does not include the number of attached FAs, the random variables, $\mathbf{c}^{j+1} - \mathbf{c}^j$, for different values of j , are not independent. For example, if there are 2 FAs and given some nonzero value for the random variable, $\mathbf{c}^{j+1} - \mathbf{c}^j$, within an interval that would satisfy the state of going from none attached to one attached, or from two attached to one attached, or from one attached to two attached, then that probability would be greater than if it was conditioned on the previous random variable being 0.

Numerical simulations were conducted to see how Eq (4) compares with the experimental MSD as a function of τ . The x's in Fig 7 are calculated using Eqs (4), (13) and (7) in the following way. Since our process is space and time homogeneous we assume $j = 0$ and $\mathbf{c}^0 = (0, 0)$ to give

$$\mathbb{E}_\rho[\mathbf{c}^{j+1} - \mathbf{c}^j] = \mathbb{E}_\rho[\mathbf{c}^1 - \mathbf{c}^0] = \mathbb{E}_\rho[\mathbf{c}^1].$$

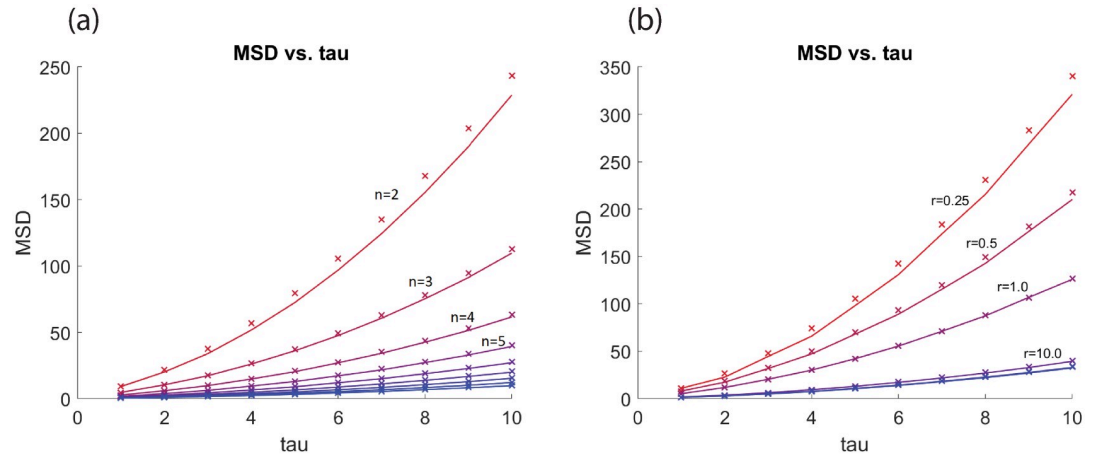


Fig 7. MSD as a function of tau. Panel (a) visualizes the numerical ($TAMSD_2$) and theoretical ($AMSD$) results of the MSD versus tau for different values of n (2–10), the total number of FAs, and $r = 10$. Panel (b) shows simulations where r varies from 0.25, 0.5, 1, 10, 50, and 100 with $n = 5$. The lines represent the relationship between tau and $TAMSD_2$. The “x” uses the $AMSD$ (Eq (13)) and expectation (Eq (7)) to compute the MSD in Eq (4) for different values of tau. There were 1,000,000 iterations.

<https://doi.org/10.1371/journal.pone.0261021.g007>

This relates Eq (7) to the expectation term in Eq (4). Furthermore,

$$MSD(1) = \mathbb{E}_\rho[\|\mathbf{c}^1 - \mathbf{c}^0\|^2] = \mathbf{Var}_\rho[\|\mathbf{c}^1\|] + \|\mathbb{E}_\rho[\mathbf{c}^1]\|^2.$$

Thus using $AMSD$ from Eq (13) and the expectation found in Eq (7) to compute the variance and expectation,

$$\begin{aligned} & \tau \cdot \mathbf{Var}_\rho[\|\mathbf{c}^1\|] + \tau^2 \cdot \|\mathbb{E}_\rho[\mathbf{c}^1]\|^2 \\ & \approx \tau(AMSD(1) - \|\mathbb{E}_\rho[\mathbf{c}^1]\|^2) + \tau^2 \|\mathbb{E}_\rho[\mathbf{c}^1]\|^2 \\ & = \frac{\tau \mathbb{E}_v[\|\boldsymbol{\eta}\|^2]}{2(1+r)^{n-1}} \left(1 + \sum_{k=1}^{n-1} \binom{n-1}{k} \frac{r^k}{(k+1)^2} + \binom{n-1}{k} \frac{r^k}{(k+1)(k^2)} \sum_{i=1}^k \frac{i}{i+1} \right) \\ & + \frac{(\tau^2 - \tau) \|\mathbb{E}_v[\boldsymbol{\eta}]\|^2}{4(1+r)^{2(n-1)}} \left(1 + \sum_{k=1}^n \frac{\left(r^{k-1}(1-r) \binom{n-1}{k-1} + r^k \binom{n}{k} \right) r(n-k)}{(k+r(n-k))(k+1)} \right)^2. \end{aligned}$$

This gives an estimate of the MSD as a function of τ as indicated by Eq (4) with \mathbf{c} being the random variable. In Fig 7 panel (a) the different curves represent different values of n the total number of FAs and panel (b) shows curves where r , the propensity to attach, varies.

The graphs reveal that adhesiveness of the cell may be inferred by examining how quickly the MSD increases with respect to τ . As the number of FAs increase (or the adhesive ability of the cell), the MSD curve flattens. The same is true for the propensity of FAs to attach. As the value of r increases the MSD curve flattens. This work shows that additional information about the cell and its motility can be gained by examining the MSD.

Discussion on the MSD estimate

In this section we give an explanation as to why our estimate for the MSD is so close to the MSD calculated from simulations. In the section on the lower bound, we found the probability for being in a sequential configuration of k attachments when the history and the creation

story coincide denoted by $P_k^d(r)$ and given in Eq (14). For each k , $1 \leq k \leq n$, $\lim_{r \rightarrow \infty} P_k^d(r) = 0$. The total probability of being in this particular sequential configuration for any number of attachments and then detaching is the sum over all k of P_k^d , so as r increases sufficiently, the probability, $\sum_{k=1}^n P_k^d$ decreases (approaching 0). For $k = 1$, $\lim_{r \rightarrow 0} P_k^d(r) = .5$, but for $2 \leq k \leq n$, $\lim_{r \rightarrow 0} P_k^d(r) = 0$. Again, the total probability of being in this particular sequential configuration for any number of attachments and then detaching is the sum over all k of P_k^d , so as r decreases sufficiently, the probability $\sum_{k=1}^n P_k^d$ increases (approaching .5). Because of the bounds on the number of FAs, over a long enough simulation, on average, the probability of being in a state of any number of attachments and then detaching and the probability of being in a state of any number of attachments and then attaching is equal, and is .5. This helped us better understand why Eq (13) is such a good estimate for the MSD. Heuristically, as r decreases sufficiently, the number of attachments decreases, and the sequential probability increases, implying that the random variable (RV) values, $c^{j+1} - c^j$, being used for a detachment event (Eqs (9) and (10)) are closer to the actual values of the RVs. As r increases sufficiently, the number of attachments on average approaches the total number of FAs. Because of the initial condition of starting the simulation with no attachments, FAs quickly attach (r is large) until most are attached and the system stays in a highly attached state. Because the majority of attachments happened quickly they will be close to a sequential configuration. Thus the RVs being used for a detachment event (Eqs (9) and (11)) are still a good estimate for the MSD. For the “middle” values of r , the estimate is not as good, but is still adequate.

Conclusion

MSD is a measure of the overall drift of a particle and can be a useful tool for understanding cell motion because it also indicates mode of transport. We introduced a mathematical model for cell motion and discussed it in the context of a centroid model (a discrete-time jump process). We were able to find a good estimate for the theoretical MSD of the centroid model by introducing the concept of a sequential configuration. We found the displacement after an attach event and the displacement after a detach event when in a sequential configuration. Using the displacement for sequential configurations to approximate all detach events we found a good approximation for the MSD with a delay of one event. To further quantify the MSD, we found a lower and upper bound for the MSD. We surmised that the estimate for the MSD had a small relative error because the FA configuration frequently is in a sequential configuration or close to it.

Supporting information

S1 File.
(PDF)

Author Contributions

Conceptualization: Christopher P. Grant, J. C. Dallon.

Formal analysis: Mary Ellen Rosen, Christopher P. Grant, J. C. Dallon.

Investigation: Mary Ellen Rosen.

Methodology: Mary Ellen Rosen, Christopher P. Grant, J. C. Dallon.

Software: Mary Ellen Rosen, J. C. Dallon.

Supervision: Christopher P. Grant, J. C. Dallon.

Validation: Mary Ellen Rosen, Christopher P. Grant, J. C. Dallon.

Visualization: Mary Ellen Rosen, J. C. Dallon.

Writing – original draft: Mary Ellen Rosen, J. C. Dallon.

Writing – review & editing: Mary Ellen Rosen, Christopher P. Grant, J. C. Dallon.

References

1. Bretscher MS. How Animal Cells Move. *Scientific American*. 1987; 257(6):72–93. <https://doi.org/10.1038/scientificamerican1287-72> PMID: 3452926
2. Qian H, Sheetz MP, Elson EL. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical Journal*. 1991; 60(4):910–921. [https://doi.org/10.1016/S0006-3495\(91\)82125-7](https://doi.org/10.1016/S0006-3495(91)82125-7) PMID: 1742458
3. Sikora G, Teuerle M, Wylomańska A, Grebenkov D. Statistical properties of the anomalous scaling exponent estimator based on time-averaged mean-square displacement. *Physical Review E*. 2017; 96(2):022132. <https://doi.org/10.1103/PhysRevE.96.022132> PMID: 28950534
4. Rebenshtok A, Barkai E. Distribution of Time-Averaged Observables for Weak Ergodicity Breaking. *Physical Review Letters*. 2007; 99(21):210601. <https://doi.org/10.1103/PhysRevLett.99.210601> PMID: 18233203
5. Mason TG, Ganesan K, van Zanten JH, Wirtz D, Kuo SC. Particle Tracking Microrheology of Complex Fluids. *Physical Review Letters*. 1997; 79(17):3282–3285. <https://doi.org/10.1103/PhysRevLett.79.3282>
6. Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*. 1905; 322(8):549–560. <https://doi.org/10.1002/andp.19053220806>
7. Squires TM, Brady JF. A simple paradigm for active and nonlinear microrheology. *Physics of Fluids*. 2005; 17(7):073101. <https://doi.org/10.1063/1.1960607>
8. Sutherland W. LXXV. A dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1905; 9(54):781–785. <https://doi.org/10.1080/14786440509463331>
9. Dasgupta BR, Tee SY, Crocker JC, Frisken BJ, Weitz DA. Microrheology of polyethylene oxide using diffusing wave spectroscopy and single scattering. *Physical Review E*. 2002; 65(5):051505. <https://doi.org/10.1103/PhysRevE.65.051505> PMID: 12059562
10. Xu J, Viasnoff V, Wirtz D. Compliance of actin filament networks measured by particle-tracking microrheology and diffusing wave spectroscopy. *Rheologica Acta*. 1998; 37(4):387–398. <https://doi.org/10.1007/s003970050125>
11. Saxton MJ, Jacobson K. SINGLE-PARTICLE TRACKING: Applications to Membrane Dynamics. *Annual Review of Biophysics and Biomolecular Structure*. 1997; 26(1):373–399. <https://doi.org/10.1146/annurev.biophys.26.1.373> PMID: 9241424
12. Weihs D, Mason TG, Teitell MA. Bio-Microrheology: A Frontier in Microrheology. *Biophysical Journal*. 2006; 91(11):4296–4305. <https://doi.org/10.1529/biophysj.106.081109> PMID: 16963507
13. Gal N, Lechtman-Goldstein D, Weihs D. Particle tracking in living cells: a review of the mean square displacement method and beyond. *Rheologica Acta*. 2013; 52(5):425–443. <https://doi.org/10.1007/s00397-013-0694-6>
14. Burov S, Jeon JH, Metzler R, Barkai E. Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking. *Physical Chemistry Chemical Physics*. 2011; 13(5):1800–1812. <https://doi.org/10.1039/c0cp01879a> PMID: 21203639
15. Reverey JF, Jeon JH, Bao H, Leippe M, Metzler R, Selhuber-Unkel C. Superdiffusion dominates intracellular particle motion in the supercrowded cytoplasm of pathogenic *Acanthamoeba castellanii*. *Scientific Reports*. 2015; 5(1):1–14. <https://doi.org/10.1038/srep11690> PMID: 26123798
16. Hoffman BD, Massiera G, Citters KMV, Crocker JC. The consensus mechanics of cultured mammalian cells. *Proceedings of the National Academy of Sciences*. 2006; 103(27):10259–10264. <https://doi.org/10.1073/pnas.0510348103> PMID: 16793927

17. Gal N, Weihs D. Intracellular Mechanics and Activity of Breast Cancer Cells Correlate with Metastatic Potential. *Cell Biochemistry and Biophysics*. 2012; 63(3):199–209. <https://doi.org/10.1007/s12013-012-9356-z> PMID: 22555560
18. Pearson K. The Problem of the Random Walk. *Nature*. 1905; 72(1867):342–342. <https://doi.org/10.1038/072342a0>
19. Codling Edward A, Plank Michael J, Benhamou Simon. Random walk models in biology. *Journal of The Royal Society Interface*. 2008; 5(25):813–834. <https://doi.org/10.1098/rsif.2008.0014> PMID: 18426776
20. Grimmett G, Grimmett GR, Grimmett PoMSG, Stirzaker D, Stirzaker MIDR. *Probability and Random Processes*. OUP Oxford; 2001.
21. Dallon JC, Scott M, Smith WV. A force based model of individual cell migration with discrete attachment sites and random switching terms. *Journal of Biomechanical Engineering*. 2013; 135(7):71008. <https://doi.org/10.1115/1.4023987> PMID: 23722520
22. Dallon JC, Evans EJ, Grant CP, Smith WV. Cell speed is independent of force in a mathematical model of amoeboidal cell motion with random switching terms. *Mathematical Biosciences*. 2013; 246(1):1–7. <https://doi.org/10.1016/j.mbs.2013.09.005> PMID: 24060706