4 Diffusion, random walks, fluctuation-dissipation theorem

4.1 Motivation

- Diffusion is a very important transport phenomenon: it can move matter, energy (heat), information etc. It is not the only transport mechanism, however: the alternative is transport by flows. We will try to characterize the difference mathematically. Intuitively, in flow substance travels with some velocity (units of [distance]/[time]), i.e. the flow is directed. In diffusion, the substance spreads with some diffusion constant (units of [distance]$^2$/[time]), with no particular direction of spread.

- Diffusion is not so important for transport of matter on “everyday” length scales (meters) and time scales (seconds, minutes, hours). For instance, mixing in oceans and atmosphere are all done through flows (e.g. convection because of temperature differences). Also consider the example of the spread of odor from a little flask in the corner of the room, with $D \sim 10 - 100 \text{ mm}^2/\text{s}$ and across a distance of $L = 10 \text{ m}$. If you estimate the time (as in the Fermi problem), $T \sim L^2/D \sim 1 \cdot 10^6 \text{ s}$ (which is around 10 days), it is clear that smell doesn’t spread by diffusion either at these distances. In contrast, diffusion is critical in the microscopic world, on scales from nanometers to millimeters. Indeed, we implicitly assume that diffusion is so efficient that it mixes up spatially the chemical reactants very fast, whenever we write down the classical reaction-kinetics equations (for instance, for enzymatic reactions). There is no spatial dependence in those reactions! If space were important (i.e. important that it takes some time for one molecule to come from one end of the cell to the other by diffusion), we would need to write down reaction-diffusion equations instead, which are partial differential equations containing both space and time. Some characteristic scales for the microscopic world: diffusion of proteins in the cell (depends very much on their size) $\sim 1 - 50 \text{ } \mu \text{m}^2/\text{s}$, diffusion of small molecules, several $100 \text{ } \mu \text{m}^2/\text{s}$, of transcription factors $\sim 0.1 - 10 \text{ } \mu \text{m}^2/\text{s}$. So for a transcription factor with $D \sim 10 \text{ } \mu \text{m}^2/\text{s}$ in a cell (or nucleus) with radius $R \sim \mu \text{m}$, the characteristic time to cross the cell is about a second. The example of Bicoid gradient establishment from an anterior source, where $L \sim 400 \text{ } \mu \text{m}$ and $T \sim 1 \text{ hour}$, which seems hardly consistent with the measured bicoid diffusion of between $0.1 - 10 \text{ } \mu \text{m}^2/\text{s}$. Perhaps there is active transport / flows?

- The mathematics of diffusion is much more general than just for describing real material transport. The same equations can be used to describe allele frequencies in population genetics, the number of stochastically produced mRNA, the price of fluctuating stock in the stock market, transfer of heat or diffusion of probability in inference problems. In short, the formalism needed to deal with diffusion is directly useful for a much broader set of things.

4.2 Random walk in 1D

Consider a 1D problem with discrete positions (discrete spacing $\Delta x$), evolving in discrete time (time bins $\Delta t$). We consider a particle on this lattice that makes, at every time step, a discrete jump by one lattice distance $\Delta x$, at at time $t = 0$ starts at position 0. The rate of going left is $k_L$, the rate of going right is $k_R$. If $\Delta t$ is small enough, then, the probability of jumping left is $p_L = k_L \Delta t$, of jumping right is $p_R = k_R \Delta t$ and the probability of staying put is $p_0 = 1 - p_L - p_R$, as required by normalization.
We want to know where the particle is going to be after time $t$, on average. Averaging is over repeated trials over this process, since there is a randomness in the sequence of concrete left/right steps that my random walker makes. Let’s start by writing down the equation:

$$x_t = x_{t-\Delta t} + (\Delta x)\xi_t,$$  \hspace{1cm} (1)

where $\xi_t$ is a random variable, that takes the value $+1$ (rightward step) with probability $p_L$, value $-1$ (leftward step) with $p_L$, and value 0 (stay put) with $p_0$. The expected position after time $t$ is then (by taking the expectation of the above equation):

$$\langle x_t \rangle = \langle x_{t-\Delta t} \rangle + (\Delta x)\langle \xi_t \rangle.$$ \hspace{1cm} (2)

Note that we can recursively apply the same equation to get $\langle x_{t-\Delta t} \rangle$, until we get:

$$\langle x_t \rangle = \langle x_0 \rangle + (\Delta x)\left[\langle \xi_t \rangle + \langle \xi_{t-\Delta t} \rangle + \ldots\right].$$ \hspace{1cm} (3)

Now, $\langle x_0 \rangle = 0$, since this is the initial condition. The parenthesis contain $t/\Delta t$ terms in the sum, because this is the total number of steps during time $t$. Importantly, since the steps are all independent and the probabilities of stepping left/right do not depend on time or position, each of these averages is equal to the rest. We can evaluate:

$$\langle \xi \rangle = (+1)p_R + (-1)p_L + (0)p_0 = p_R - p_L,$$ \hspace{1cm} (4)

by the definition of the expectation value of a random variable. Putting everything together, we find:

$$\langle x_t \rangle = t \left(\frac{\Delta x}{\Delta t}\right) (p_R - p_L) = \tilde{v}t.$$ \hspace{1cm} (5)

That is, the average displacement is proportional to time, and increases with the “flow velocity” $\tilde{v} = \Delta x/\Delta t(p_R - p_L)$. This is a characteristic of the directed flow.

Similarly, we can ask about the spread (variance) around that mean position, and can compute it equivalently. We look for variance after time $t$, or $\sigma_t^2 = \langle x_t^2 \rangle - \langle x_t \rangle^2$. The second term, the square of the expectation, we already know from Eq. (5). To compute the first term, first square equation of motion and then take the expectation, to get:

$$\langle x_t^2 \rangle = \langle (x_{t-\Delta t})^2 \rangle + (\Delta x)^2\langle \xi_t^2 \rangle + 2(\Delta x)\langle x_{t-\Delta t}\rangle\langle \xi_t \rangle.$$ \hspace{1cm} (6)

We can evaluate:

$$\langle \xi^2 \rangle = (+1)^2p_R + (-1)^2p_L + (0)^2p_0 = p_R + p_L,$$ \hspace{1cm} (7)

Using these equations together and using the recursive trick again, show that $\sigma_t^2 = \langle x_t^2 \rangle - \langle x_t \rangle^2$ can be written as:

$$\sigma_t^2 = 2\tilde{D}t,$$ \hspace{1cm} (8)

where $\tilde{D} = \frac{1}{2}\frac{(\Delta x)^2}{\Delta t}(p_R + p_L - (p_R - p_L)^2)$. Here the variance of position grows linearly with time, with “diffusion coefficient” $\tilde{D}$. That is, the displacement grows as a square root of time, which is a characteristic of a diffusion process. Although we have shown this in a 1D simple case, this is generally true in arbitrary dimension, and is conceptually at the core of diffusion.

Note that if $p_R = p_L$, there are no flows, and one only has diffusive spread. In contrast, if $p_L = 1$ (or $p_R = 1$), there is no diffusion (check that $\tilde{D} = 0$), since the random walker is not random anymore: it only hops (and flows) left or only right.
In short, in the same system we see two distinct and fundamental modes of transport: flow and diffusion, that differ in how displacement grows with time.

If you plot displacement vs time (increasing linearly with flow and as a square root for diffusion), you will see that the two curves intersect. For displacements smaller than the point of intersection, diffusion is effective—i.e. faster than flow—as a transport mode; for displacement larger than that point, flow is more effective. Quantitatively, if I have object of size $L$, to cross it by diffusion takes $\sim L^2/D$ time, and to cross it by flow takes $L/v$, where $v$ is the flow velocity. For a given size, the times to cross distance $L$ are comparable when $L \sim D/v$. If the size is smaller than the ratio $L \ll D/v$ (or equivalently $Pe = Lv/D \ll 1$, where $Pe$ is a dimensionless number known as the Peclet number), then for a given object of size $L$ diffusion is more efficient as a transport mechanism. For $L \gg D/v$ (or $Pe \gg 1$), flow is more efficient. This is crucial for biology: transport by diffusion is “free”, driven by random thermal motion. To maintain flows, you always have to pay energy, because flows are nonequilibrium phenomena that exist only when you maintain gradients. For example, in circulatory systems you need a pump (a heart) to maintain a pressure gradient that supports the blood flow. We need the circulatory system because we are too large to exchange nutrients and oxygen across our bodies by diffusion alone.

4.3 Three examples where diffusion places physical—and therefore derivable—limits on the biological systems

- Transport networks in animals. Established observation that metabolic rate, $R$, scales with the mass of an animal, $m$, with a power law that has an approximately $3/4$ slope, i.e. $R \propto m^{3/4}$, for several orders of magnitude in the animal size. $R$ is taken to be the rate at which some crucial nutrient is consumed (say oxygen), and thus the flow of that nutrient into the circulatory system. The proposed explanation is that this scaling emerges very generally from various transport architectures (like a branching hierarchical transport system), because (i) most of the properties of such systems, like the branching pattern of the network, scale (are self-similar), but (ii) the ultimate volume unit of the tissue, at the end of the transport system, where capillaries reach, is independent of the size of the animal. This is because ultimately the tissues of all animals are similar, and the transport from capillaries to the cells close-by all organisms use the same kind of diffusive transport. So the range of tissue that can feed itself by diffusion from the capillary does NOT scale with the size of the animal. Putting these things together yields the observed $3/4$ scaling (c.f. works by JH Brown et al, Banavar et al). Interestingly, while single cells and small organisms solve their supply problems by diffusion alone, large multicellular organisms extended their transport range by putting a flow network on top of what is essentially still diffusive transport at small scale (from capillaries to cells), because at those spatial scale diffusion is effective.

- Diffusion-limited association rate. Suppose you have two reactant types, A and B, and when they meet they react to form a compound. Historically an interesting problem has been to ask what is the FASTEST forward reaction rate. The progress could be made when it was realized that there are two steps to this reaction: first, the reactants have to meet by chance while diffusing around in solution; second, while they are close, they need to react. The first contribution to the total reaction time is due to purely physical (transport) reasons, quite independent from the molecular details in A and B. The second contribution is due to chemical reasons, and can vary a lot with the chemical details. Since the total time for the reaction is the sum of these contributions, the rates add in inverse:
where the total forward rate of reaction \( k_{\text{tot}} \) is decomposed into the diffusive \( k_d \) and reaction \( k_r \) parts. Now the MAXIMUM achievable total rate is limited by the diffusion rate, even if the reaction rate were infinitely fast. Thus \( k_{\text{max}} = k_d = 4\pi Da \) (known as the Smoluchowski limit), where \( D \) is the diffusion constant of the reactants and \( a \) is their molecular size (physical dimension). Again, here we have the case where there is a basic limit, containing only the fundamental physics and depending only on basic properties, like the molecular size, to how quickly a (potentially complicated) reaction can occur. Note that the reaction can be an association of a transcription factor (molecule A) to its binding site (molecule B). The total maximal reaction rate for this association is \( k_{\text{tot}}c \), where \( c \) is the concentration of transcription factor, and \( k_{\text{tot}} \), in the Smoluchowski limit, can be computed by estimating \( D \sim 10 \text{ } \mu \text{m}^2/\text{s} \) and the binding site size \( a \sim 3 \text{ } \text{nm} \). The typical TF concentrations are in the nanomolar range; convince yourself that \( 1 \text{nmol} \leq c \leq 1 \text{mM} \).

Putting everything together, the total rate of TF, present in nanomolar concentration, to a single binding site, is \( k_{\text{max}}c \approx 0.2 \text{s}^{-1} \), i.e. the time to find the binding site is on the order of one to ten seconds. The Smoluchowski limit can approximately be derived by the following argument: suppose molecules of A at concentration \( c \) are trying to find a target molecule B (of which, for simplicity, let there be only one) of linear dimension \( a \). Let’s consider the volume \( V \) around the target, of radius \( R \) such that in that radius there is one molecule of A on average; thus \( Vc \sim 1 \) and hence \( R^3 \sim 3/(4\pi c) \). This molecule of A has a probability of \( a^3/R^3 \) of hitting (and reacting with) the target every time it samples a small volume of space of linear size \( a \). To sample another independent volume element of space it has to move, by diffusion, by at least \( a \), which takes \( \tau \sim a^2/D \). The rate at which molecules of A react with the target is \( k_+c \), where \( k_+ \) is the diffusion limited association rate we are looking for; \( k_+c \) is also the probability per unit time of A molecules finding the target. Per attempt, the probability is \( a^3/R^3 \) as we argued before, and each attempt lasts for time \( \tau \); thus \( k_+c \sim a^3/R^3/\tau \sim 4\pi Da c/3 \), giving \( k_+ \sim 4\pi Da /3 \) in this very rough estimation (cf. with the exact result, \( k_+ = 4\pi Da /3 \)).

- **Berg-Purcell limit to precision in chemical detection.** See the homework problem.

### 4.4 Diffusion equation and the mathematics of diffusion

We started with a discrete time / discrete space random walk:

\[
x_t = x_{t-\Delta t} + \Delta x \xi_t
\]

where \( \langle \xi_t \rangle = p_R - p_L \) and \( \langle \xi_t^2 \rangle = p_R + p_L \). Can we make the limit where time becomes continuous? Note we can rewrite the above equation as:

\[
(x_t - x_{t-\Delta t})/\Delta t = \xi_t \Delta x / \Delta t
\]

As I make \( \Delta t \) very small, the left hand side converges to \( dx/dt \). On the right hand side I now have a new random process, whose mean rate is equal to \( (p_R - p_L) \Delta x / \Delta t \), but remember that from our definition of rates that is just \( (k_R - k_L) \Delta x \) (which has units of velocity, as it should). The new process also “produces variance,” at the rate \( (k_R + k_L) \Delta x^2 / \Delta t \). This is a bit harder to think about, because this formally diverges when \( \Delta t \to 0 \). I can rewrite the continuous limit in the traditional way as

\[
\frac{dx}{dt} = v + \eta(t),
\]

\[\text{(11)}\]
where the average rate (flow or drift part $v$) has been pulled out of the random variable, and the remaining random part, $\eta(t)$ is now zero mean, $\langle \eta(t) \rangle = 0$, and again produces the variance at the specified rate $(k_R + k_L) \Delta x^2 / \Delta t$. Mathematically we commonly write this as:
\[
\langle \eta(t) \eta(t') \rangle = 2D \delta(t - t'),
\]
introducing a diffusion parameter $D$. $\delta(t - t')$ is known as the Dirac-delta function, which is infinite when $t = t'$ and zero otherwise (and it integrates to 1, i.e., $\int_{-\infty}^{\infty} dt \delta(t - t_0) = 1$). Note that $C(t, t') = \langle \eta(t) \eta(t') \rangle$ is a correlation function (as we learned before), and since it is zero EXCEPT when $t = t'$, we are saying that this random force $\eta$ is uncorrelated. When $t = t'$, the correlation function is just the variance, and this is what we computed in the case of discrete random walker (i.e. the variance of the random step). While this random variable $\eta$ seems very unnatural, and has a somewhat funny correlation function (a Dirac-delta function), its power spectrum—as we learned the last time—is perfectly well defined and indeed extremely simple: it is just a constant as a function of frequency. So going to Fourier space helps a lot here.

In general, continuous differential equations of this form, where there is a “random” component (or random force) are known as Langevin equations in physics, and as stochastic differential equations (SDE) or Ito equations in math.

There is a very different way (with equivalent results) to analyze diffusion. Instead of writing down an equation for the dynamical variable ($x$) which needs to have random force (making the equation Langevin / SDE), we write down the differential equation governing the probability that we observe position $x$ at given time $t$. This is an equivalent description using partial differential equations.

To derive an equation for $P(x|t)$ (the probability of observing the random walker at position $x$ at time $t$), we first realize that we have already computed the mean of this distribution, $\langle x_t \rangle$, and the variance, $\sigma^2_t$, using our SDE. Here we like to have the equation for the full probability distribution.

To derive this, let’s look at why $P(x|t)$ can change. First, because the particle can jump left (with rate $k_L$) or right (with $k_R$) the probability of staying at $x$ will decrease with time. But it can also increase because the particle can jump one cell right from position $x - \Delta x$ (and land at $x$), or one cell left from position $x + \Delta x$ (and also land at $x$). Putting together all these contributions, we find:
\[
\frac{\partial P(x|t)}{\partial t} = -(k_L + k_R)P(x|t) + k_R P(x - \Delta x|t) + k_L P(x + \Delta x|t),
\]
We can now Taylor expand $P(x \pm \Delta x|t) = P(x|t) \pm \frac{\partial P(x|t)}{\partial x} \Delta x + \frac{1}{2} \frac{\partial^2 P(x|t)}{\partial x^2} \Delta x^2$, and insert back into the equation (since we assumed my spatial step is 1 for simplicity). After collecting all terms together, we get:
\[
\frac{\partial P(x|t)}{\partial t} = -(k_R - k_L) \Delta x \frac{\partial P(x|t)}{\partial x} + \frac{1}{2} (k_R + k_L) \Delta x^2 \frac{\partial^2 P(x|t)}{\partial x^2},
\]
a partial differential equation that governs the probability of observing random walker at $x$ after time $t$, given that there is flow (first term) and random diffusion (second term).

In general, note the correspondence of the two descriptions (where we write in general $v$ for the rate of flow, and $D$ for the rate of diffusion, which in the case of 1D random walker we worked out in terms of $k_L, k_R$):
\[
\frac{\partial P(x|t)}{\partial t} = -v \frac{\partial P(x|t)}{\partial x} + D \frac{\partial^2 P(x|t)}{\partial x^2},
\]
\[
\frac{dx}{dt} = v + \eta(t),
\]
where $\langle \eta(t) \rangle = 0$, $\langle \eta(t) \eta(t') \rangle = 2D \delta(t - t')$. So you can either describe the same process with the probabilistic PDE equation (known as diffusion-advection equation or the Fokker Planck equation), or with the kinetic SDE / Langevin equation. Note that both equations contain the flow part (velocity $v$) and the random part (diffusion $D$ and the variance of the random force $\eta$). You can always jump from one to the other description, as is necessary.

Traditionally, you might recognize the diffusion equation easily if you get rid of the flow part, and instead of talking about the probability talk about the concentration of particles, $c(x,t)$ (ultimately, the probability of finding a particle at a given location and the concentration of particles differ just in a factor, and since the equation is linear, this factor cancels out):

$$\frac{\partial c(x,t)}{\partial t} = D \frac{\partial^2 c(x,t)}{\partial x^2}$$

(17)

This is the well-known diffusion equation. For the simplest case of the random walker starting at $x = 0$ at $t = 0$, or in the continuous case, concentration at $t = 0$ being infinitely high at $x = x_0$ and 0 elsewhere ($c(x,t = 0) = \delta(x - x_0)$, where $\delta$ is again our Dirac-delta function), we can ask what is the solution to this equation. Think about the random walker first: it makes many independent random steps, and we are interested in the sum of all such steps, to find the distribution of final position $x$ after time $t$. By central limit theorem, the distribution of final positions has to be closer and closer to a Gaussian, the more steps the walker makes. Equivalently, we could guess that the solution to the diffusion equation with the initial condition $c(x,t = 0) = \delta(x)$ is also Gaussian, $G(x|t,x_0)$:

$$G(x|t,x_0) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{(x-x_0)^2}{4Dt}}.$$  

(18)

You may verify that this is indeed the solution of the Eq. (17) by inserting it in the equation and seeing that the left- and right-hand sides are equal. Thus a drop of ink in a container of water—if there is really no mixing flows—spreads as a gaussian whose variance grows with time as $\sigma^2 = 2Dt$, precisely equivalent to what we concluded in the first part of the lecture. This is the basic law of diffusion. If you look in 2 or 3 dimensions, the law remains valid, only the factor changes: $\sigma^2 = 4Dt$ in two dimensions, and $6Dt$ in three.

The solution of the problem for point initial condition (as here) is called the Green’s function or the fundamental solution. If you start with some distribution of ink which is not concentrated at one point, you can always write it as a sum of such point sources. And since diffusion equation is linear ($c$ enters as a first power only in the equation), if the initial condition is a sum of point sources, the result is the sum of results for point sources, so if I start with some distribution of ink $\rho(x_0)$, the solution is:

$$P(x|t) = \int \rho(x_0)G(x|t,x_0)dx_0.$$  

(19)

### 4.5 Conclusion

- Diffusion is a particular kind of transport driven by random, undirected motion or mixing, where distance grows with a square root of time. Since it is random thermal motion, it can proceed without paying energy cost to support (unlike the directed flows).

- Diffusion can be an efficient transport / mixing mechanism over short distances. This is quantitatively captured by the Peclet number, $Pe$, of the system (compare $Pe$ to 1 to see if flows or diffusion are more efficient).
• Since diffusion is such a general physical phenomenon that often precedes other more complicated chemical processes, it can put physical limits to the speed or precision of those processes that are independent of the fine molecular details of what’s going on, but depend only on basic physical properties like the size (e.g. of the molecules or the system), diffusion constant, etc.

• The mathematics of diffusion is much more broadly applicable than just to material diffusion. Heat diffusion is governed by the same equation, allele frequency change as well. 1D discrete random walk can be a model for generating single molecules in chemical reactions (where jump left is degrading one molecule and jump right is creating one molecule, each happening at given rates). Stock fluctuations are often described by SDE equations. There is two different-looking but formally equivalent ways to describe random walk / diffusion mathematically. One is by writing down an equation for dynamical variable $x$ directly (Langevin / SDE equation), the other for probability of observing $x$ (Diffusion / Fokker-Planck PDE equation). Physicists in general call equations of the type

$$\frac{\partial}{\partial t}(\cdot) = -v \frac{\partial}{\partial x}(\cdot) + D \frac{\partial^2}{\partial x^2}(\cdot)$$

Fokker-Planck equations (if they are for probability of something), or diffusion equations, as long as they have the mathematical structure above and irrespective of what is the actual quantity denoted by $(\cdot)$. The common part is the evolution due combined influence of directed part (flow) and undirected spread (diffusion).