

# Probabilistic descriptions of biological phenomena

Hyun Youk

(Dated: January 16, 2019)

In this lecture, we discuss the basic, mathematical properties of probabilities and probabilistic description of two biological settings - (1) diffusion of (bio-)molecules and (2) the limits to how accurately cells can sense concentrations of molecules.

## I. PROBABILITIES

The goal of this section is to introduce or review the basic properties of probabilities. We will start with a simple example - rolling a six-sided dice - and then generalize our observations.

### A. Discrete probabilities

Let's start with a simple example. Suppose you roll a six-sided dice. If it's a fair dice (i.e., the dice is not engineered to land more on one face than any other face), then the chance of getting any one face is  $1/6$  (we say "one in six" chance). We say that the **probability** of getting a particular face on the dice is  $1/6$ . To formalize this, we can say that  $X$  is the **outcome value**. Then we have:  $X = 1, 2, 3, 4, 5, \text{ or } 6$ . Then, mathematically, we write

$$P(6) = \frac{1}{6} \quad (1)$$

to mean that the probability of getting a "6" after throwing the dice is  $1/6$ . Motivated by this definition, we can formulate a general definition of probability.

**Definition of probability:** Suppose that there are  $N$  possible outcomes in total ( $N \geq 1$ ). Let  $X$  be the outcome value. Then, we define the **probability of getting an outcome of value of  $X$**  to be

$$P(X) = \frac{\text{number of ways of getting the value } X}{\text{total number of possible outcomes}} \quad (2)$$

Equivalently, we can also say

$$P(X) = \frac{\text{number of ways of getting the **desired outcome-value } X}{\text{total number of possible outcomes}} \quad (3)**$$

According to above definition of probabilities, the probability of getting a particular outcome-value  $X$  must be less than or equal to 1 since there cannot be more ways of getting the desired value  $X$  than the total number of possible outcome-values. Moreover, the probability of getting a particular outcome-value  $X$  must be larger than or equal to 0 since either there is at least one way of getting an outcome-value  $X$  or none at all. So, we must have

$$0 \leq P(X) \leq 1 \quad (4)$$

**Question:** Suppose we now throw the dice twice. What is the probability of getting a "2" in the first time and then "3" in the second time?

**Answer:** To answer this question, we use the general definition of probability (Eq. 2 and, equivalently, Eq. 3). Let  $x$  be the outcome value for the first dice-throw and  $y$  be the outcome value for the second dice-throw. Then,  $(x, y)$  summarizes the outcome. The total number  $N$  of possible outcomes (i.e., the total number of possible values  $(x, y)$ ) is  $N = 6 \times 6 = 36$  since there are six possible values of  $x$  and six possible values of  $y$ . Out of these, only one outcome,  $(2, 3)$ , is the desired outcome. Thus, we write

$$P(2 \text{ and } 3) = \frac{1}{36} \quad (5)$$

to denote the probability of getting (2, 3). We can also rewrite above as

$$P(2 \text{ and } 3) = \frac{1}{6} \cdot \frac{1}{6} \quad (6a)$$

$$= P(2) \cdot P(3) \quad (6b)$$

Above is no accident. The reason is that, by definition,

$$P((x, y)) = \frac{(\text{number of ways of getting } x) \cdot (\text{number of ways of getting } y)}{\text{total number of outcomes}} \quad (7a)$$

$$= \frac{(\text{number of ways of getting } x) \cdot (\text{number of ways of getting } y)}{(\text{total \# of outcomes for the first dice-throw}) \cdot (\text{total \# of outcomes for second dice-throw})} \quad (7b)$$

$$= \left( \frac{\text{number of ways of getting } x}{\text{total \# of outcomes for the first dice-throw}} \right) \cdot \left( \frac{\text{number of ways of getting } y}{\text{total \# of outcomes for the second dice-throw}} \right) \quad (7c)$$

$$= P(x) \cdot P(y) \quad (7d)$$

Motivated by this example, we can generalize as follows: First, we say "event" or "experiment" to refer to a process that yields an outcome value, such as throwing a dice. Another example of an event or an experiment is blindly picking a card from a deck of Poker cards. Moreover, suppose that we have two **events** that are **independent of each other** like in case of throwing of the dice twice - that is, the outcome of one event does not depend on the other events. If  $X$  is the outcome value of the first experiment and  $Y$  is the outcome value of the second experiment, then by above derivation, we can see that

$$P(X, Y) = P(X) \cdot P(Y) \quad (8)$$

We say that Eq. 8 is the **joint probability** for two independent events. In fact, we can generalize to  $M$  independent events. Using the derivation given in Eqs. 7a - 7d, it follows that if  $X_1, X_2, \dots, X_M$  are the outcome values of each of the  $M$  independent experiments, then

$$P(X_1, X_2, \dots, X_M) = P(X_1) \cdot P(X_2) \cdot \dots \cdot P(X_M) \quad (9)$$

**Question:** Suppose we now throw the six-side dice once. What is the probability of getting *either* a "2" or a "5"?

**Answer:** Here, our desired outcome is getting either a "2" or a "5". We would be happy with either one. In this case, by the definition of probability (Eq. 3), we have

$$P(2 \text{ or } 5) = \frac{\# \text{ of ways of getting the desired outcome}}{\text{total \# of possible outcomes}} \quad (10a)$$

$$= \frac{2}{6} \quad (10b)$$

Note that we can rewrite above as

$$P(2 \text{ or } 5) = \frac{\# \text{ of ways of getting the desired outcome}}{\text{total } \# \text{ of possible outcomes}} \quad (11a)$$

$$= \frac{(\# \text{ of ways of getting a "2"}) + (\# \text{ of ways of getting a "5"})}{\text{total } \# \text{ of possible outcomes}} \quad (11b)$$

$$= \frac{\# \text{ of ways of getting a "2"}}{\text{total } \# \text{ of possible outcomes}} + \frac{\# \text{ of ways of getting a "5"}}{\text{total } \# \text{ of possible outcomes}} \quad (11c)$$

$$= P(2) + P(5) \quad (11d)$$

$$= \frac{1}{6} + \frac{1}{6} \quad (11e)$$

$$= \frac{2}{6} \quad (11f)$$

Motivated by this example, we can generalize. For two events, the probability of getting outcome values of **either**  $X$  **or**  $Y$  is

$$P(X \text{ or } Y) = P(X) + P(Y) \quad (12)$$

We can further to say that the probability of getting one of  $M$  desired outcome values (denoted  $X_1, X_2, \dots, X_M$ ) out of a total of  $N$  possible outcomes in one experiment is

$$P(X_1 \text{ or } X_2 \text{ or } \dots \text{ or } X_M) = P(X_1) + P(X_2) + \dots + P(X_M) \quad (13)$$

**Normalization condition:** The final property of probability that we discuss, before introducing statistical quantities that are based on probabilities, is called the **normalization condition**. This simply says that if we add up the probabilities of each outcome, we should get 1. That is, if there is a set of  $N$  possible, distinct outcome-values for an experiment, denoted by  $\{x_1, x_2, \dots, x_N\}$ , then

$$P(x_1) + P(x_2) + \dots + P(x_N) = 1 \quad (14)$$

This makes sense since above equation is just a rewrite of Eq. 13 with  $M = N$  - it is the probability that we get one of the possible values after an experiment (in other words, the probability that we get any value that an experiment can generate after performing that experiment):

$$P(x_1 \text{ or } x_2 \text{ or } \dots \text{ or } x_N) = P(x_1) + P(x_2) + \dots + P(x_N) \quad (15a)$$

$$= 1 \quad (15b)$$

We can compactly write Eq. 14 as

$$\sum_{i=1}^N P(x_i) = 1 \quad (16)$$

In summary, summing up the probabilities for each possible outcome should yield 1.

### Definitions of statistical quantities:

Now that we know how to calculate probabilities of events, let's define five key **statistical quantities** - (1) **random variable**, (2) **average** (also called the mean or the **expectation value**), (3) **standard deviation**, (4) **variance**, and (5) **fractional error**.

**(1) Random variable:** A random variable  $X$  is a variable whose value is the outcome value of an experiment. Thus, there is a probability  $P(x_i)$  of  $X$  taking on a particular value  $x_i$ . If there is a total of  $N$  possible outcome values ( $x_1, x_2, \dots, x_N$ ), then  $X$  can take on any one of the  $N$  values. So we only know probabilistically, not exactly, the value of  $X$ . In the example of throwing a six-sided dice, the outcome-value is the random variable  $X$ , which can take on one of  $\{1, 2, 3, 4, 5, 6\}$  as a value.

Note that a sum of two random variables is also a random variable. That is, if  $X$  and  $Y$  are random variables, then so is  $X + Y$  because the summed value is also probabilistic - we cannot definitely predict the summed value before the experiment of measuring  $X$  and  $Y$  because we are unsure of the value of  $X$  and the value of  $Y$  before the experiment. For the same reason, a definite constant (e.g., 3) times a random variable is also a random variable (e.g.,  $3X$ ). Likewise, product of two random variables,  $XY$ , is also a random variable.

**(2) Mean:** The mean value (i.e., **expectation value, average**) of a random variable  $X$  is written as  $\langle X \rangle$  and defined as

$$\langle X \rangle = x_1P(x_1) + x_2P(x_2) + \dots + x_NP(x_N) \quad (17)$$

where  $\{x_1, x_2, x_3, \dots, x_N\}$  is the set of  $N$  possible, distinct outcome-values and  $P(x_i)$  is the probability of getting the outcome-value  $x_i$ . Applying this definition to a throwing of the six-sided dice, the expectation value of the outcome is

$$\langle X \rangle = 1P(1) + 2P(2) + \dots + 6P(6) \quad (18a)$$

$$= \frac{1}{6} + \frac{2}{6} + \dots + \frac{6}{6} \quad (18b)$$

$$= 3.5 \quad (18c)$$

which matches our expectation (thus the term, "expectation value"). We can write Eq. 17 more compactly as

$$\langle X \rangle = \sum_{i=1}^N x_iP(x_i) \quad (19)$$

Eq. 19 is the definition of the mean-value of random variable  $X$ . There are two properties of the mean that we will use. The first deals with the mean of a new random variable that we form by multiplying a random variable  $X$  by a constant  $c$ . The resulting, new random-variable  $cX$  - note that this is random since we cannot definitely predict its value before doing an experiment - has the following mean:

$$\langle cX \rangle = \sum_{i=1}^N cx_iP(x_i) \quad (20a)$$

$$= c \sum_{i=1}^N x_iP(x_i) \quad (20b)$$

$$= c \langle X \rangle \quad (20c)$$

The second property of the mean deals with the mean of a new random-variable that we form by adding a constant  $c$  to a random variable  $X$ . The resulting, new random-variable  $(c + X)$  is also a random variable for the same reason that  $cX$  is a random variable. Its mean is

$$\langle c + X \rangle = \sum_{i=1}^N (c + x_i)P(x_i) \quad (21a)$$

$$= c \sum_{i=1}^N P(x_i) + \sum_{i=1}^N x_iP(x_i) \quad (21b)$$

$$= c + \langle X \rangle \quad (21c)$$

where we used the normalization condition (Eq. 16) in the last line.

**(3) Standard deviation:** The standard deviation  $\sigma$  (Greek letter "sigma") of a random variable  $X$  is defined as

$$\sigma = \sqrt{\langle (X - \langle X \rangle)^2 \rangle} \quad (22)$$

The standard deviation (Eq. 22) quantifies the expected deviation of a random variable's value from the mean. You might wonder why the square and the square root are in Eq. 22. You might say that a more natural way measuring the expected (mean) deviation of a random variable's value from its mean would be

$$\langle X - \langle X \rangle \rangle \quad (\text{why not this as the definition of } \sigma?) \quad (23)$$

But this would not work. In fact, for *any* random variable, we have

$$\langle X - \langle X \rangle \rangle = 0 \quad (24)$$

and thus this  $\langle X - \langle X \rangle \rangle$  is not an informative quantity for what we want. To see why Eq. 24 is true, note that since  $\langle X \rangle$  is a constant, Eq. 21c tells us that

$$\langle X - \langle X \rangle \rangle = \langle X \rangle - \langle X \rangle = 0 \quad (25)$$

Eq. 25 also reveals a deeper meaning of the expectation value  $\langle X \rangle$ : The mean  $\langle X \rangle$  is the value for which the random variable  $X$  fluctuates, on average, just as many times above  $\langle X \rangle$  (i.e.,  $(X - \langle X \rangle) > 0$ ) as it does below  $\langle X \rangle$  (i.e.,  $(X - \langle X \rangle) < 0$ ), which leads to  $\langle X - \langle X \rangle \rangle = 0$ . By defining the standard deviation as in Eq. 22, we make all values of  $(X - \langle X \rangle)$  to be positive or zero. We can make sense of the standard deviation (Eq. 22) as follows:

$$(X - \langle X \rangle)^2 \quad (\text{square of the distance between } X \text{ and } \langle X \rangle) \quad (26)$$

and then

$$\langle (X - \langle X \rangle)^2 \rangle \quad (\text{average of the square of the distance between } X \text{ and } \langle X \rangle) \quad (27)$$

and thus

$$\sigma = \sqrt{\langle (X - \langle X \rangle)^2 \rangle} \quad (\text{average "distance" between } X \text{ and } \langle X \rangle; \text{ square root "takes away" the square}) \quad (28)$$

note that above equation says "distance" with the quotes since  $\sigma$  is not exactly the average distance (the true average distance is zero (by Eq. 25)).

There is a convenient way to calculate the standard deviation. Note that

$$\sigma = \sqrt{\langle (X - \langle X \rangle)^2 \rangle} \quad (29a)$$

$$= \sqrt{\langle X^2 - 2X\langle X \rangle - \langle X \rangle^2 \rangle} \quad (29b)$$

$$= \sqrt{\langle X^2 \rangle - 2\langle X \rangle\langle X \rangle - \langle X \rangle^2} \quad (\text{by Eqs. 20c and 21c}) \quad (29c)$$

$$= \sqrt{\langle X^2 \rangle - \langle X \rangle^2} \quad (\text{convenient way to calculate } \sigma) \quad (29d)$$

**(4) Variance:** The variance is defined as  $\sigma^2$ :

$$\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle \quad (30)$$

The variance is useful because not having the square root in the standard deviation (Eq. 22) makes calculations simpler sometimes. In light of Eq. 29d, we have

$$\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2 \quad (31)$$

**(5) Fractional error:** The fractional error is defined as

$$\frac{\sigma}{\langle X \rangle} = \frac{\sqrt{\langle (X - \langle X \rangle)^2 \rangle}}{\langle X \rangle} \quad (32)$$

The fractional error is useful for determining how large the standard deviation is compared to the mean (i.e., as a percentage of the mean). Although not important for our purpose, there is also a related quantity called the **Coefficient of Variation (CV)**:

$$\text{CV} = \frac{\sigma^2}{\langle X \rangle} \quad (33)$$

## B. Continuous probabilities

We can define all the quantities above but now for continuous probabilities. The previous section dealt with **discrete probabilities**. There, we could count the total number of outcomes,  $N$ , which was a positive integer (i.e.,  $N = 1, 2, 3, \dots$ ). This is because the random variable (the outcome) took on discrete values, such as in the case of rolling a six-sided dice. **Continuous probabilities** for a random variable describe experiments whose outcome takes on a **continuum of values**. Here are two such examples.

**Example A:** With limited information, you use a microscope to observe a single cell. Can you predict at what time  $t$  after you begin observing the cell, it will divide?

**Answer:** Here, the time  $t$  is a random variable because you cannot exactly predict it with the limited information given to you (in fact, in real experiments, we never have enough information to predict typical behaviours of cells such as its division time  $t$ ). But we can determine the probability that a cell divides time  $t$  after you begin observing it. Note that  $t$  can be any real number. That is, it can be any real number in the range,  $0 < t < \infty$ . So  $t$  can be 3.1415, or 3.14159, or 3141592, or 4.1, or 5, or 5.01, or 5.001, or 5.0001, and so on. You get the point. There is an infinite number of possible values for  $t$ . Of course, some values of  $t$  are less likely than others. For instance, intuition tells you that for a fast-dividing bacterial cell,  $t$  being a million years is very unlikely (and thus probability for that should be nearly, if not exactly, zero).

**Example B:** A point-sized particle is confined between two walls, one at  $x = 0$  and the other at  $x = L$  ( $L$  is the distance between the walls). It bounces back and forth between the two walls without loss of speed and it only moves along the x-axis only. While the particle is moving, you close your eyes and then you open them. Can you predict its position  $x$  before you open your eyes?

**Answer:** Here, the particle's position  $x$  is a random variable because you cannot definitely predict it without looking. But unlike in the previous example with  $t$ , the  $x$  is confined within a finite range:  $0 \leq x \leq L$  ( $t$  was not confined to a finite range because  $t$  could be arbitrarily large). But  $x$  can still take on a continuum of values and, in fact, there is still an infinite amount of values of  $x$ . For example,  $x = L/2$ ,  $x = L/2 + 0.1 * L$ ,  $x = L/2 + 0.01L$ ,  $x = L/2 + 0.011L$ ,  $x = L/2 + 0.0112L$ , and so on. You get the point.

We just saw two examples of random variables whose values lie in a continuum. We now want to describe the probability for each outcome and then extend the definitions of statistical quantities so that they are defined for both discrete and continuous random variables. First, let's take example B and ask, "what is the probability that you find the particle between  $x = 0$  and  $x = L/2$  when you open your eyes?". According to our intuition, the probability should be  $1/2$  (and it indeed is, as our calculation will show). We can also intuitively see, without any calculations,

that the probability of finding the particle between  $x = L/2$  and  $x = L$  is also  $1/2$  because the particle is moving back-and-forth without preferring one half of the room over the other. Note that

$$\frac{1}{2} = \frac{L/2}{L} \quad (34)$$

Following our intuition, the probability of  $x$  being  $0 < x < L/4$  should be  $1/4$  as is the probability of it being  $L/4 < x < 2L/4$ . Like wise, the probability of finding the particle between  $x = 3L/4$  and  $x = L$  should be  $1/4$ . Note that

$$\frac{1}{4} = \frac{L/4}{L} \quad (35)$$

We can see a pattern here. The probability of finding the particle between  $x_0$  and  $x_0 + \Delta x$  ( $\Delta x > 0$ ) is

$$P((x_0, x_0 + \Delta x)) = \frac{\Delta x}{L} \quad (36)$$

where  $P((x, x + \Delta x))$  denotes the probability of the random variable (position)  $x$  being within the interval  $(x_0, x_0 + \Delta x)$  for *any* value of  $x_0$  (assuming that  $x_0$  is properly sized so that  $x_0 + \Delta x \leq L$ ). According to Eq. 36, the probability of finding the particle *exactly* at position  $x_0$  is

$$P((x_0, x_0 + 0)) = \frac{0}{L} = 0 \quad (37)$$

no matter what the value of  $x_0$  is. This makes intuitive sense. It is simply saying that there are so many (in fact, infinitely many) values that  $x$  can take on within the range  $[0, L]$  that the probability of  $x$  being *exactly* equal to  $x_0$  (to 130 decimal places, if  $x_0$  has exactly 130 decimal places), is zero. In other words, according to the definition of probability (Eq. 2),

$$P(x = x_0) = \frac{1}{N} = \frac{1}{\infty} = 0 \quad (38)$$

Note that the definition of probability of an outcome (Eq. 2) applies to both discrete and continuous random variables.

A convenient way to express Eq. 36 is by defining a **Probability Density Function (PDF)**  $\rho(x_0)$  (Greek letter "rho"):

$$\rho(x_0) = \frac{1}{L} \quad (\text{Probability Density Function (PDF) for Example B}) \quad (39)$$

which then lets us write Eq. 36 as

$$P((x_0, x_0 + \Delta x)) = \rho(x_0)\Delta x \quad (40)$$

Motivated by this example, we can generalize the concept of PDF to any situation, not just to a particle confined between two walls. For any situation, we define **probability density function (PDF)** for a random variable  $y$  to be  $\rho(y)$  so that

$$P((y_0, y_0 + dy)) = \rho(y_0)dy \quad (\text{General definition of Probability Density Function (PDF)}) \quad (41)$$

where  $P((y_0, y_0 + dy))$  is the probability of the random variable  $y$  having a value in the infinitesimal interval  $(y_0, y_0 + dy)$  ( $y_0$  is some particular value). Note that, in general,  $\rho(y)$  does not have to be a constant function. Some values of  $y$  may be more probable than others. As we will see, certain functional forms of  $\rho(y)$  are given special names, such as **Poisson distribution**, **normal distribution**, and **uniform distribution**. For a finite interval  $(y_0, y_0 + \Delta y)$ , in which  $\Delta y$  is not infinitesimal like  $dy$ , Eq. 41 tells us that

$$P((y_0, y_0 + \Delta y)) = \int_{y_0}^{y_0 + \Delta y} \rho(y) dy \quad (\text{How to use a PDF}) \quad (42)$$

where  $P((y_0, y_0 + \Delta y))$  is the probability of the random variable  $y$  being a value within (a potentially large) interval  $(y_0, y_0 + \Delta y)$ .

**Normalization condition:** The normalization condition (Eq. 16) for continuous PDF is

$$1 = \int_{y_{\min}}^{y_{\max}} \rho(y) dy \quad (43)$$

where  $y_{\min}$  and  $y_{\max}$  are minimum and maximum possible values of the random variable  $y$ . Note that they can be  $\pm\infty$ .

**Definitions of statistical quantities:**

For the most part, the definitions for statistical quantities that we gave for the discrete random variable are exactly the same for continuous random variables. But we repeat them here for completeness.

(1) **Mean:** The expectation value of a random variable  $y$ , given a continuous PDF  $\rho(y)$ , is

$$\langle y \rangle = \int_{y_{\min}}^{y_{\max}} y \rho(y) dy \quad (44)$$

where  $y_{\min}$  and  $y_{\max}$  are minimum and maximum possible values of the random variable  $y$ .

(2) **Standard deviation  $\sigma$ :** Same as in the case of discrete random variables:

$$\sigma = \sqrt{\langle y^2 \rangle - \langle y \rangle^2} \quad (45)$$

(3) **Variance  $\sigma^2$ :** Same as in the case of discrete random variables:

$$\sigma^2 = \langle y^2 \rangle - \langle y \rangle^2 \quad (46)$$

The rest (fractional error and CV) are also exactly the same for both discrete and continuous random variables. We see above that the only difference is in how we calculate the mean for the two kinds of random variables (actually, when you learn more math (e.g., Dirac delta function), you will see that even the mean is computed in the same way for both discrete and continuous random variables - i.e. Eq. 44 and Eq. 19 are identical).

## II. PROBABILISTIC (STOCHASTIC) DESCRIPTION OF BIOLOGICAL SYSTEMS

It is now time to use the mathematics of probabilities that we discussed above to biological systems. What we are about to do falls in the intersection of **statistical physics** (i.e., using probabilities to describe atoms and molecules) and **quantitative biology** (i.e., using math to describe biological systems). We will apply this to two settings: (1) **diffusion (Brownian motion) of "objects"** (including biomolecules and randomly walking cells), and (2) **Berg-Purcell limit**, which is the limit to how accurately a cell or any receptor can determine the concentration of an external molecule.

### A. Diffusion and random walk

In this section, we introduce the concept of **diffusion** and analyze the one-dimensional diffusion of objects (molecule, cell, randomly walking person). Eventually, we will derive the one-dimensional **diffusion equation**. Molecules inside



cells move by diffusion. Moreover, the mathematics of diffusion are important because they are applied to study a wide range of biological phenomena, not just diffusing bio-molecules, such as stochastic gene-expression (i.e., mRNA and proteins are randomly transcribed and translated respectively).

Suppose that we have a molecule that is constrained to move along a line (the x-axis). The molecule moves in a special manner. It takes one step, of length  $L$ , after every time interval  $\delta t$ . Although its step-size is fixed at  $L$ , it randomly chooses to which direction - right or left - it will move for each step. For this reason, we say that this molecule takes a **random walk**. Although not required for our purpose, let's make our analysis simpler by assuming that it is equally likely to move to the right as it does to the left (i.e., probability is  $1/2$  for both directions).

**Question 1:** What is the mean displacement after  $N$  steps (i.e., after time  $N\delta t$ ) if the molecule starts from  $x = 0$ ?  
**Solution:** There are multiple ways to address this. One quick way is to note that this is like flipping a coin  $N$  times. For a regular coin, probability is  $1/2$  for getting a 'head' as is for getting a 'tail'. Replacing the 'head' with a 'step to the right' and the 'tail' with a 'step to the left' maps the coin-tossing problem to the molecule's random walk. Then, we expect that the average position after  $N$  steps,  $\langle x_N \rangle$ , should not be any more to the right of the molecule's starting position ( $x = 0$ ) as it is to the left of it. In other words,

$$\langle x_N \rangle = 0 \quad (\text{for any } N \geq 0) \quad (47)$$

Thus, the average displacement is:  $\langle x_N - x_0 \rangle = \langle x_N \rangle = 0$ . Note that when  $N = 0$ , we trivially have  $\langle x_0 \rangle = 0$  since the molecule starts at  $x = 0$ . We can also get the same result through an honest calculation as follows. First, let's denote  $\langle x_n \rangle$  to be the average position after  $n$  steps ( $n \geq 0$ ). Then,

$$\langle x_1 \rangle = \frac{L}{2} + \frac{-L}{2} = 0 \quad (\text{by the definition of the mean: Eq. 19}) \quad (48)$$

And thus,

$$\begin{aligned} \langle x_2 \rangle &= \langle x_1 + aL \rangle && (\text{where } a = \pm 1 \text{ is a random variable}) && (49a) \\ &= \langle x_1 \rangle + \langle a \rangle L && (\text{by Eqs. 21c and 20c}) && (49b) \\ &= \langle x_1 \rangle && (\langle a \rangle = 0 \text{ since } a \text{ is equally likely to be } +1 \text{ as } -1) && (49c) \\ &= 0 && (\text{by Eq. 48}) && (49d) \end{aligned}$$

In fact, we can now see that

$$\langle x_N \rangle = \langle x_{N-1} \rangle + \underbrace{\langle a \rangle}_{=0} L \quad (50a)$$

$$= \langle x_{N-1} \rangle \quad (50b)$$

$$= \langle x_{N-2} \rangle + \underbrace{\langle a \rangle}_{=0} L \quad (50c)$$

$$= \langle x_{N-2} \rangle \quad (50d)$$

$$\vdots \quad (50e)$$

$$= \langle x_0 \rangle \quad (50f)$$

$$= 0 \quad (50g)$$

In other words,  $\langle x_N \rangle = 0$  for all  $N \geq 0$  and hence the molecule's mean displacement after  $N$  steps is

$$\langle x_N \rangle - x_0 = 0 \quad (51)$$

as our intuition had told us.

We can more easily interpret above result by imagining that we perform an experiment with randomly walking people instead of molecules. Suppose that we have 1000 people and that each person takes  $N$  random walks with the

rule that everyone takes the same step-length  $L$  for every step and flips a fair coin before each step to decide whether to move to the right or left, thus ensuring that the probability for each direction is  $1/2$ . Then, Eq. 51 says that after time  $N\delta t$ , everyone has taken exactly  $N$  steps (since everyone takes one step for every time interval  $\delta t$ ) and that if we average the position of everyone, we would get  $x = 0$ . But this does *not* mean that all 1000 people are standing exactly at  $x = 0$  after taking  $N$  steps. Indeed, our intuition tells us - assuming that people can pass by each other on the line without blocking each other - that if we take a picture of where everyone is after beginning the experiment, we would find some *spread* in people's positions. Eq. 51 simply tells us that we should expect to find just as many people to the left of  $x = 0$  as there are to the right of  $x = 0$ . From our definitions of statistical quantities, we can see that this spread is exactly what the **standard deviation in displacement after  $N$  steps**,  $\sigma(N)$  measures (see Eq. 22). That leads us to our next question.

**Question 2:** What is the standard deviation in displacement after  $N$  steps?

**Answer:** We can calculate  $\sigma(N)$  if we calculate the variance,  $\sigma^2(N)$ , through the relationship, Eq. 29d, as follows:

$$\sigma^2(N) = \langle (x_N - x_0)^2 \rangle - \langle x_N - x_0 \rangle^2 \quad (\text{by Eq. 29d}) \quad (52a)$$

$$= \langle x_N^2 \rangle - \langle x_N \rangle^2 \quad (\text{since } x_0 = 0) \quad (52b)$$

$$= \langle x_N^2 \rangle \quad (\text{since } \langle x_N \rangle = 0) \quad (52c)$$

Thus, we just need to calculate the mean value of  $x_N^2$ . We can do this by first noting that:

$$\langle x_N^2 \rangle = \langle (x_{N-1} + aL)^2 \rangle \quad (\text{where } a = \pm 1 \text{ is a random variable}) \quad (53a)$$

$$= \langle x_{N-1}^2 + 2aLx_{N-1} + a^2L^2 \rangle \quad (53b)$$

$$= \langle x_{N-1}^2 \rangle + 2L \langle ax_{N-1} \rangle + L^2 \langle a^2 \rangle \quad (\text{using Eqs. 20c and 21c}) \quad (53c)$$

$$= \langle x_{N-1}^2 \rangle + L^2 \langle a^2 \rangle \quad (\text{since } \langle ax_{N-1} \rangle = \langle \pm x_{N-1} \rangle = 0) \quad (53d)$$

$$= \langle x_{N-1}^2 \rangle + L^2 \quad (\text{since } \langle a^2 \rangle = \langle 1 \rangle = 1) \quad (53e)$$

Since above is true for any integer  $N$  ( $N \geq 1$ ), we obtain the following **recursion relationship**:

$$\langle x_N^2 \rangle = \langle x_{N-1}^2 \rangle + L^2 \quad (54a)$$

$$= \langle x_{N-2}^2 \rangle + L^2 + L^2 \quad (54b)$$

$$= \langle x_{N-3}^2 \rangle + L^2 + L^2 + L^2 \quad (54c)$$

$$\vdots \quad (54d)$$

$$= \langle x_0^2 \rangle + \underbrace{L^2 + L^2 + \dots + L^2}_{N \text{ times}} \quad (54e)$$

$$= NL^2 \quad (\text{since } \langle x_0^2 \rangle = 0) \quad (54f)$$

for  $N \geq 1$ . So by Eq. 52c, we have the **variance in in displacement after  $N$  steps**  $\sigma^2(N)$ ,

$$\sigma^2(N) = NL^2 \quad (55)$$

Now, we can convert the  $N$  into time  $t$ . Since everyone (molecules) take one step for each time interval  $\delta t$ , we have  $N = t/\delta t$  and hence

$$\sigma^2(t) = \frac{tL^2}{\delta t} \quad (56)$$

By defining a new quantity  $D$  that we call the **diffusion constant** as

$$D = \frac{L^2}{2\delta t} \quad (\text{definition of the diffusion constant } D) \quad (57)$$

we can rewrite Eq. 56 as

$$\sigma^2(t) = \langle x^2(t) \rangle = 2Dt \quad (\text{mean-squared displacement after time } t) \quad (58)$$

Note that the diffusion constant has units of length<sup>2</sup>/time. Eq. 58 is a fundamental result, which we derived from pure logic, is the famous equation called **mean-squared displacement** (since  $\sigma^2(N) = \langle x_N^2 \rangle$ ). The standard deviation in the position of the diffusing molecules, after they all start from position  $x = 0$  at  $t = 0$ , is thus

$$\sigma(t) = \sqrt{2Dt} \quad (59)$$

A small caveat here is that we assumed that  $t/\delta t$  is exactly an integer  $N$ . That does not have to be the case. But by making  $\delta t$  to be infinitesimal, we achieve two things: (1) the molecule really is more like a molecule - it is continuously moving, not stopping and then taking a step after some finite time interval, and (2)  $t/\delta t$  being an integer is becomes a better and better approximation as  $\delta t$  becomes smaller and smaller. In this case, we interpret  $L$  as the typical distance that the molecule travels in a straight line before it switches its direction. The variance (Eq. 58) is more useful than and is often instead of the standard deviation (Eq. 59). Experimentally, one can measure the diffusion constant  $D$  by releasing many molecules from one location (e.g., use a pipette tip to inject many copies of a molecule at a small location in a liquid) and then measuring the spread in the positions (variance) after time  $t$ . In such an experiment, the molecules (e.g., dye molecule) would diffuse in three dimensions, not the 1D diffusion that we examined to derive Eq. 58. This leads us to the next question.

**Question 3:** What is the mean-squared displacement of molecules diffusing in two- or three-dimensions?

**Answer:** Your biophysics course will show you the derivation. We will just state the results without derivations for now. The mean-squared displacement after time  $t$  for molecules diffusing in two dimensions is

$$\sigma_{2D}^2(t) = \langle |\vec{r}(t)|^2 \rangle = 4Dt \quad (\text{mean-squared displacement for 2D diffusion}) \quad (60)$$

and in three-dimensions is

$$\sigma_{3D}^2(t) = \langle |\vec{r}(t)|^2 \rangle = 6Dt \quad (\text{mean-squared displacement for 3D diffusion}) \quad (61)$$

where  $\vec{r}(t)$  is the position of the molecule at time  $t$ . Looking at Eqs. 58, 60, and 61, we see the following pattern:

$$\sigma_{nD}^2(t) = \langle |\vec{r}(t)|^2 \rangle = 2nDt \quad (\text{mean-squared displacement for } n\text{-dimensional diffusion}) \quad (62)$$

The derivation of the mean-squared displacement in 2D and 3D diffusions are actually nearly identical to that of the 1D diffusion that we performed, as you will learn in a biophysics course.

Finally, you might worry that we derived  $\sigma^2(t)$  (Eq. 58) under a very special circumstance - the molecule took an equal step size every time and had an equal probability for stepping to the right as it did for the left - and thus Eq. 58 may not hold for more general situations, such as when a molecule takes steps of different sizes. This leads us to our next and final question.

**Question 4:** Does the result for  $\sigma^2(t)$  (Eq. 58) still hold for more general 1D-diffusions such as molecules whose step sizes are not fixed and are determined by complicated probabilities?

**Answer:** Yes. You will show this in a problem set.

## B. Berg-Purcell limit: Physics limits how accurately cells can sense their environment

In this section, we summarize a classic and one of biophysicists' favourite papers - "[Physics of chemoreception](#)" by [Howard Berg](#) and [Edward Purcell](#) in *Biophysical Journal* (1977). Here, Berg and Purcell asked and addressed whether there is a fundamental physical limit to how accurately a cell can detect the concentration of a diffusing molecule. They discovered, from first-principles calculations, that there is such a limit and this limit applies not just to living cells but to all non-living detectors, including any detection device that one may conceive of in the future. Experimentally, Howard Berg and others have verified this limit in the bacterium *E. coli* that senses concentration of

desired molecules and swims towards it. This lower bound is also useful in understanding how accurately cells inside a developing embryo can sense a concentration of a **morphogen** - a chemical that cues an embryonic cell what type of a specialized cell it should differentiate into - and thus in understanding how accurately an embryo develops into a fully-formed organism (e.g., fly embryo developing into a fly). In this section, we will derive this limit.

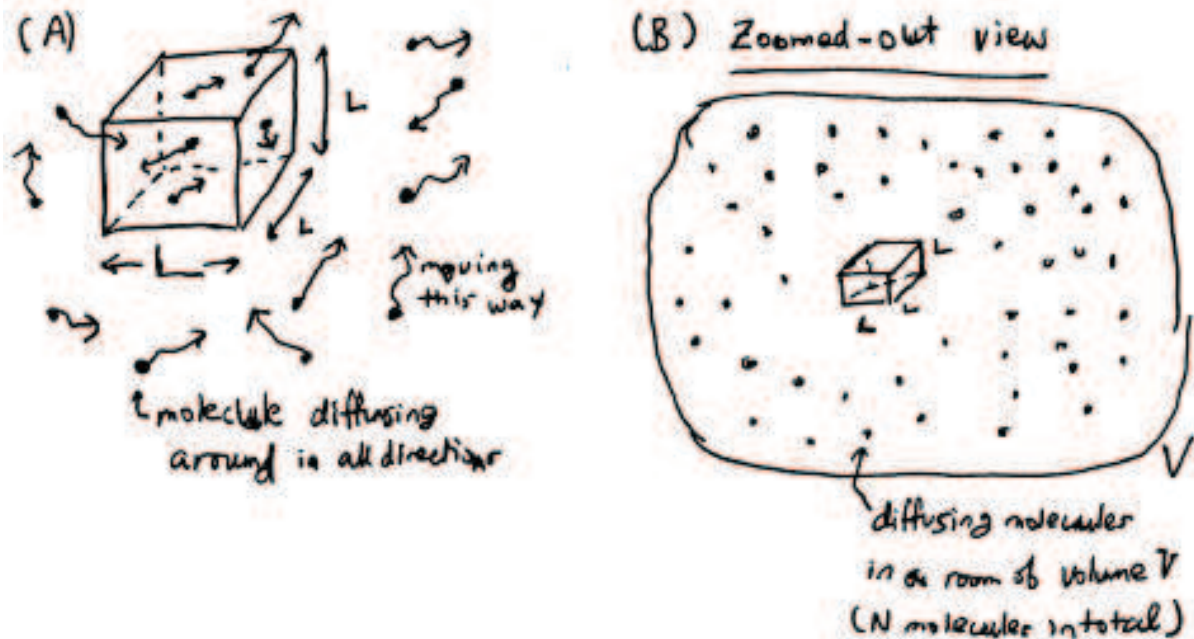


Figure 1. Setup for deriving the Berg-Purcell limit:  $N$  identical molecules are diffusing around in a large room of volume  $V$ . Inside the room is a cubic detector whose sides have length  $L$  and whose walls are completely permeable to the molecules. The detector "measures" how many molecules are inside it at a given time.

Consider a cubic detector with a side length  $L$  (Fig. 1). This detector could be the entire cell, or a receptor inside a cell, or a location on DNA where a transcription factor should bind (then  $L$  is the length of that portion of DNA), or a receptor on the cell surface, just to list some of the many possibilities. The detector sits inside a much larger "bath" of volume  $V$  (i.e.,  $V \gg L^3$ ) (Fig. 1). Suppose that in this large bath,  $N$  molecules are diffusing around with a diffusion constant  $D$ . Let's assume that the  $N$  molecules are uniformly distributed inside the large bath. The average concentration  $\langle c \rangle$  is then

$$\langle c \rangle = \frac{N}{V} \quad (63)$$

The main issue is that while we know the exact average concentration because we already *see* the  $N$  molecules inside the box in one snapshot, the cell does *not* know the total number of molecules inside the box. In fact, the cell may not encounter every one of the  $N$  molecules in a given amount of time because the molecules are diffusing (randomly moving) inside the box and thus some molecules may not *hit* the cell in a given time interval. The cell uses its detector to measure ("count") the number of molecules that come inside the detector. In this way, the cell samples a sub region in the large bath with its cubic detector, count the number of molecules that are inside the detector, and then from this deduces the concentration. On average, the number of molecules  $\langle n \rangle$  inside the cubic detector is

$$\langle n \rangle = \langle c \rangle L^3 \quad (64)$$

But this is the average value. If the cell measures the number of molecules  $x$  times, it will not obtain exactly the same value  $x$  times. Let's calculate how much variability there will be in the number  $n$  (random variable) among the different measurements. Let's consider one particular molecule. If we take a snapshot of the system, the probability  $p$  that we will find this particular molecule inside the cubic volume is

$$p = \frac{L^3}{V} \quad (65)$$

This is similar in logic to Example B above (the particle trapped between two walls). Now, suppose we paint a number on every one of the  $N$  molecules. So one molecule will have number "1" painted on it. The next one will have "2" on it. And so on. Now, out of these numbered molecules, let's consider a particular set of molecules -  $\{1, 2, \dots, n\}$  (where  $1 \leq n \leq N$ ). We now ask what the probability of finding this set of molecules within the detector is. Since each molecule is diffusing independently of each other, the probability of one molecule to be in the box is independent of what any of the other molecules are doing. Thus,

$$\underbrace{p \cdot p \cdot \dots \cdot p}_{n \text{ times}} = p^n \quad (\text{using Eq. 9}) \quad (66)$$

is the probability of finding the particular set of numbered molecules,  $\{1, 2, \dots, n\}$ , in the detector. But this calculation also includes the possibility that the other molecules, to which we assigned a number larger than  $n$ , may also be present in the detector. We actually want a probability that only the molecules in the set  $\{1, 2, \dots, n\}$  is in the detector and, at the same time, the other  $N - n$  molecules being outside the detector. In our snapshot, the probability that the particular molecule is not inside the detector is  $1-p$ . So, applying Eq. 9 and combining with Eq. 67, we have that

$$p^n(1-p)^{N-n} \quad (67)$$

is the probability of having only the set of numbered molecules,  $\{1, 2, \dots, n\}$ , being in the detector and no one else. Now, there is nothing special about the set  $\{1, 2, \dots, n\}$ . If we pick a different set of  $n$  numbered molecules, say  $\{2, 3, 4, \dots, n, n+1\}$  or  $\{N, N-1, N-2, \dots, N-n+1\}$ , the probability of having exactly that particular set of molecules in the detector and no one else would still be Eq. 67. So, the probability  $P(n)$  that exactly  $n$  molecules are inside the detector (which can be *any set* of  $n$  molecules that we pick from the  $N$  molecules) is

$$P(n) = \underbrace{p^n(1-p)^{N-n} + p^n(1-p)^{N-n} + \dots + p^n(1-p)^{N-n}}_{\text{total number of groups of } n \text{ molecules}} \quad (68a)$$

$$= g \cdot p^n(1-p)^{N-n} \quad (68b)$$

where  $g$  is the total number of ways to generate groups of  $n$  molecules from a pool of  $N$  molecules. As we will see,

$$g = \binom{N}{n} \quad (69a)$$

$$= \frac{N!}{(N-n)!n!} \quad (69b)$$

which is called the **binomial coefficient**. Here,  $k!$  (read "k factorial") means

$$k! = k(k-1)(k-2)\dots 1 \quad (\text{where } k \geq 1) \quad (70a)$$

$$\text{and } 0! = 1 \quad (0! \text{ is specially defined like this}) \quad (70b)$$

If you have not encountered the binomial coefficient before, note that  $N!$  is the total number of ways to arrange  $N$  molecules in a line. That is, if you assume that you paint a number on each molecule, from 1 to  $N$ , then you can have line up the molecules in the order  $[1, 2, 3, \dots, N]$ , or in the order  $[2, 1, 3, 4, \dots, N]$ , or in the order  $[3, 1, 2, 4, \dots, N]$ , and so on. To count how many ways of forming the line there is, note that in the first location, we have  $N$  choices. Then after we picked the first molecule, we have  $N - 1$  choices for the second position. So, there are  $N(N - 1)$  ways to form a line of length equal to two molecules (and thus a total of  $N(N - 1)$  *distinct* lines formed by two molecules). To count how many distinct lines of length equal to three there are, the same argument yields  $N(N - 1)(N - 2)$  as the total number of such lines. Finally, there must be  $N!$  distinct lines that  $N$  molecules can form. But for our detector, we are not asking about lines, in which we care about which molecule stands in front of which other molecule. We simply want to know, how many *sets* of  $n$ -molecules can be inside the cube. This means that we count an ordered line of  $[1, 2, 3, \dots, n]$  to be the same as an ordered line of  $[n, n-1, \dots, 2, 1]$  since the same numbered molecules (1 to  $n$ ) are in the box in both cases. This is where the binomial coefficient comes in. It eliminates the kind of double counting

(overcounting) that we would have if we just counted the number of distinct ordered lines. Let's unpack Eq. ?? to understand it. First, note that

$$\frac{N!}{(N-n)!} = \frac{N(N-1)(N-2)\dots(N-n+1)(N-n)!}{(N-n)!} \quad (71a)$$

$$= N(N-1)(N-2)\dots(N-n+1) \quad (71b)$$

is the total number of ways to form a line of length equal to  $n$  with  $N$  molecules (see above paragraph if you don't understand this). But as we said before, we should treat some of these ordered lines as being the same (i.e., the cubic detector cares about sets of molecules instead of lines of molecules. Take one ordered line with  $n$  molecules. For example, let's look at  $[1, 2, 3, \dots, n]$ . There are  $n!$  ways of permuting these elements. Each permutation reorders the molecules' placement in the line but we have the same set of  $n$  molecules for each  $n!$  permutation. Take another line with  $n$  molecules. For example, let's look at  $[2, 3, \dots, n, n+1]$ . There are  $n!$  ways of permuting these elements as well, each one giving a new line. But we also have the same set of  $n$  molecules for each of these lines. From this argument, we see that we have overcounted each set of  $n$ -molecules by  $n!$  times. Thus, we must divide Eq. 71b (the total number of ordered lines of length equal to  $n$ ) by  $n!$ , which yields the binomial coefficient:

$$\frac{N!}{(N-n)!n!} = \binom{N}{n} \quad (72)$$

Thus Eq. 68b becomes

$$P(n) = \binom{N}{n} p^n (1-p)^{N-n} \quad (73)$$

where  $0 \leq n \leq N$ . Eq. 73 is called the **binomial distribution** and we say that the random variable  $n$  is **binomially distributed**. The number of molecules in the box,  $n$ , is the random variable. We can calculate the average number of molecules,  $\langle n \rangle$ , found in the detector at any snap shot in time from either the definition of the mean (Eq. 19) or, more simply, by noting that a binomial distribution describes an experiment with only two outcomes: success or failure (e.g., 'heads' or 'tail' like in a coin toss). The success probability is  $p$  and the failure probability is  $1-p$ . This is like throwing a (biased) coin with a probability  $p$  of getting a 'head' and a probability of  $1-p$  of getting a 'tail'. If we throw such a coin  $N$  times, then we expect  $Np$  to be the number of times that we get a 'head' (indeed, this makes sense if  $p = 0.5$ ). Thus, we can say that

$$\langle n \rangle = Np \quad (\text{for binomial distribution}) \quad (74)$$

The variance in  $n$ , denoted by  $\sigma_n^2$ , is

$$\sigma_n^2 = N(1-p)p \quad (75)$$

which we don't derive in this course (i.e., you can just accept it as it is for this course). We can rewrite Eq. 75 as

$$\sigma_n^2 = \langle c \rangle V(1-p)p \quad (\text{from Eq. 63}) \quad (76a)$$

$$= \langle c \rangle V \left(1 - \frac{L^3}{V}\right) \frac{L^3}{V} \quad (76b)$$

$$= \langle c \rangle L^3 \left(1 - \frac{L^3}{V}\right) \quad (76c)$$

We actually want the variance associated with the measured concentration,  $\sigma_c^2$ , instead of the variance in the measured number,  $\sigma_n^2$ . We can convert  $\sigma_n^2$  to  $\sigma_c^2$  as follows:

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 \quad (77a)$$

$$= \langle c^2 L^6 \rangle - \langle c \rangle^2 L^6 \quad (77b)$$

$$= L^6 (\langle c^2 \rangle - \langle c \rangle^2) \quad (77c)$$

$$= L^6 \sigma_c^2 \quad (77d)$$

Substituting Eq. 77d into Eq. 76c yields

$$\sigma_c^2 = \langle c \rangle \left( \frac{1}{L^3} - \frac{1}{V} \right) \quad (78a)$$

$$= \langle c \rangle \left( \frac{V - L^3}{VL^3} \right) \quad (78b)$$

Assuming that  $V \gg L^3$  (i.e., the bath has much larger volume compared to the detector), Eq. 78b becomes

$$\sigma_c^2 = \langle c \rangle \left( \frac{1 - (L/V)^3}{L^3} \right) \quad (79a)$$

$$\approx \frac{\langle c \rangle}{L^3} \quad (79b)$$

This simple formula (Eq. 79b) is the variance in the measured concentration  $c$ . In other words, if the detector makes many independent measurements of the concentration inside it and then makes a histogram of all those measured concentrations, Eq. 79b is the variance in that histogram of  $c$ . The quantity that we are really interested in is not the variance itself but the **fractional error**  $\sigma_c / \langle c \rangle$  associated with the measured concentration. We are interested in the fraction error because it tells us how "big"  $\sigma_c$  is (compared to the mean (actual) concentration  $\langle c \rangle$ ). "Big" or "small" are only meaningful in science when compared to some other quantity. The fractional error in  $c$  is

$$\frac{\sigma_c}{\langle c \rangle} = \frac{1}{\sqrt{\langle c \rangle L^3}} \quad (80)$$

This is the fractional error in the detector's measurement if the detector makes just one measurement of the concentration. But suppose that the detector makes  $M$  independent measurements of the concentration. It can then calculate the average of its  $M$  independent measurements, and from this deduce the concentration of the molecule. Intuitively, this tells you that the detector can more accurately determine the concentration. We can make this statement to be more quantitatively precise. The standard deviation  $\sigma_M$  associated with the average of  $M$  independent measurements is

$$\sigma_M = \frac{\sigma_1}{\sqrt{M}} \quad (81)$$

where  $\sigma_1$  is the standard deviation when just the detector makes just one measurement (i.e.,  $M = 1$ ). From Eqs. 80 and 81, we see that the error in the detector's average of  $M$  independent measurements is

$$\frac{\sigma_c}{\langle c \rangle} = \frac{1}{\sqrt{\langle c \rangle L^3 M}} \quad (82)$$

How do we know how many measurements the cell will make before it averages those measurements? We cannot read the cell's "mind". But we can infer it from a different quantity that we can measure in experiments. Namely, suppose that we know that the cell has to determine the concentration within time interval  $T$ . Now, the question is how large can  $M$  be so that the cell can make  $M$  independent measurements within the time interval  $T$ . Clearly, the cell wants to make  $M$  to be as large as possible within the allotted time, according to Eq. 82. The key here is that the measurements must be *independent* of each other. This means that after the detector makes one measurement, it then must wait for all the molecules inside it to escape it, then wait for new molecules to enter, and then count those molecules inside. This way, the detector does not measure the same molecules in the next measurement. To see why the detector must wait until its inside is refreshed, note that if it makes the next measurement immediately after the current measurement, then it will measure the exact same value of concentration since none of the molecules inside it had time to move. In this case, the previous and the next measurements are *not* independent of each other. So we need to calculate the time it takes for the molecules inside the detector to escape. To estimate this, we use the fact that the diffusion constant  $D$  for the molecule has dimension of length<sup>2</sup>/time. So we can roughly say that a molecule requires time  $\tau = L^2/D$  to diffuse out of the detector (note that  $\tau$  has a unit of time, so this makes sense dimension-wise). Then  $T/\tau$  is the **maximum number of independent measurements** that the cell can make. In

the next section, we will see how one can estimate  $T$  in an experiment. We can now rewrite Eq. 82, in experimentally accessible parameters, as

$$\frac{\sigma_c}{\langle c \rangle} = \frac{1}{\sqrt{\langle c \rangle LTD}} \quad \text{Berg-Purcell limit} \quad (83)$$

Eq. 83 is the famous **Berg-Purcell limit**. The meaning of Eq. 83 is that if the cell has time interval  $T$  to deduce the concentration of some molecule inside it (if the detector is inside the cell) or outside it (if the detector is on the cell membrane or if the entire cell itself is the detector), then the cell cannot determine the concentration with an accuracy higher than the fraction error stated in Eq. 83. Hence the Berg-Purcell limit is the **lower bound on accuracy** placed on the detector. It is remarkable that just by using sheer logic alone, we could deduce the fundamental limit to how accurately a cell or any sensor can measure a concentration of molecules.