AP3162D: Lecture 4 - Basic modelling frameworks for developmental biology and cell-fate decisions

Hyun Youk
Delft University of Technology
(Dated: March 15, 2018)

In this lecture, we will derive the Berg-Purcell limit and apply it to developing embryo of the fruit fly (*Drosophila melanogaster*).

I. WHY IS DEVELOPMENTAL BIOLOGY FASCINATING?

Watching a single fertilized human cell (zygote) dividing many times to form a ball of cells (morula), and then seeing it gastrulate (form a cavity inside), and then eventually develop into a fully formed baby that comes out of the mother's womb is like watching a origami that folds itself. But making a time-lapse movie of a human embryo development, starting from a single cell up until birth, is of course difficult for technical and ethical reasons. Equally spectacular movies of developing embryos, which researchers have made, are of model organisms such as the fruit fly (*Drosophila melanogaster*) and the African clawed frog (*Xenopus laevis*). In this and next lecture, we will discuss common strategies and frameworks for modelling these and other types of developing embryos.

II. THE FRUIT FLY EMBRYO AS A LABORATORY

Biology is amazing because a tiny cell can grow into human beings like yourself that think, walk, eat, and solve complex problems. Studying how a single cell forms a "full" animal is the goal of a field of biology called **developmental biology**. Researchers use animals and plants to study development. One such model organism is the fruit fly, *Drosophila melanogaster*. The mother fruit fly lays an egg. Inside this egg are thousands of nuclei that detect a **morphogen protein**, **Bicoid**, that the mother deposits at the head of the embryo. In this section, we discuss how each nucleus in the embryo reads the Bicoid concentration at its location to determine their position inside the nuclei (i.e., distance from the embryo's head from where the Bicoid diffuses). Accurately determining their positions is crucial for the fly to develop the right body parts (e.g., wings) at the right location.

A. Berg-Purcell limit for detecting concentration of diffusing molecules

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).

In this paper, 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. As we will see, this lower bound is useful in understanding what cells inside an embryo must do to reliably develop. In this section, we perform a simple calculation to derive this limit.

Consider a cubic detector with a side length L. 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., $V \gg L^3$). 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 $< c >$ is then N/V. Now, the main issue here 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 some time interval since 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 $< n >$ inside the cubic detector is

$$< n >=< c > L^3 \tag{1}$$

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 among the different measurements. If we take a snapshot of the system, the probability p that we will find a particular molecule inside the cubic detector is

$$p = \frac{L^3}{V} \tag{2}$$

In our snapshot, the probability that the particular molecule is not inside the detector is 1-p. 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 the probability $P(n)$ that exactly n molecular are inside the detector is

$$P(n) = p^n(1-p)^{N-n}\binom{N}{n} \tag{3}$$

where $0 \leq n \leq N$. Eq. 3 is the **binomial distribution**. Given a molecule, it is either inside or outside the detector. The variance σ_n^2 in n (i.e., $\sigma_n^2 =< (n- < n >)^2 >=< n^2 > - < n >^2$) for n that follows the binomial distribution is

$$\sigma_n^2 = N(1-p)p \tag{4}$$

Eq. 4 comes from computing $< n^2 >$ and $< n >$ from first principles (i.e., using Eq. 3). We can rewrite Eq. 4 as

$$\sigma_n^2 =< c > V(1-p)p \tag{5a}$$
$$=< c > V\left(1 - \frac{L^3}{V}\right)\frac{L^3}{V} \tag{5b}$$
$$=< c > L^3\left(1 - \frac{L^3 < c >}{V}\right) \tag{5c}$$

We actually want the variance (error) associated with the measured concentration, σ_c^2, instead of the variance in the measured number, σ_n^2. We convert σ_n^2 to σ_c^2 by the following:

$$\sigma_n^2 =< n^2 > - < n >^2 \tag{6a}$$
$$=< c^2L^6 > - < c >^2 L^6 \tag{6b}$$
$$= L^6(< c^2 > - < c >^2) \tag{6c}$$
$$= L^6\sigma_c^2 \tag{6d}$$

Substituting Eq. 6d into Eq. 5c yields

$$\sigma_c^2 =< c > \left(\frac{1}{L^3} - \frac{< c >}{N}\right) \tag{7a}$$
$$=< c > \left(\frac{V - L^3}{VL^3}\right) \tag{7b}$$

Assuming that $V \gg L^3$ (i.e., the bath has much larger volume compared to the detector), Eq. 7b becomes

$$\sigma_c^2 \approx \frac{< c >}{L^3} \tag{8}$$

This simple formula (Eq. 8) 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. 8 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** $\sigma_c / < c >$ associated with the measured concentration. Recall that we were also interested in the fractional error in lecture 3 (see lecture note 3). We are interested in the fraction error because it tells us how "big" σ_c is (compared to the mean (actual) concentration $< c >$. "Big" or "small" are only meaningful in science when compared to some other quantity. The fractional error in c is

$$\frac{\sigma_c}{< c >} = \frac{1}{\sqrt{< c > L^3}} \tag{9}$$

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 σ_M associated with the average of M independent measurements is

$$\sigma_M = \frac{\sigma_1}{\sqrt{M}} \tag{10}$$

where σ_1 is the standard deviation when just the detector makes just one measurement (i.e., $M = 1$). From equations (5.9) and (5.10), we see that the error in the detector?s average of M independent measurements is

$$\frac{\sigma_c}{< c >} = \frac{1}{\sqrt{< c > L^3 M}} \tag{11}$$

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. 11. 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, note that the diffusion constant D for the molecule has dimension of length^2/time. So we can roughly say that a molecule requires time $\tau = L^2/D$ to diffuse out of the detector (note that τ has a unit of time, so this makes sense dimension-wise). Then T/τ 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. 11, in experimentally accessible parameters, as

$$\frac{\sigma_c}{< c >} = \frac{1}{\sqrt{< c > LTD}} \qquad \text{Berg-Purcell limit} \tag{12}$$

Eq. 12 is the famous **Berg-Purcell limit**. The meaning of Eq. 12 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. 12. Hence the Berg-Purcell limit is the **lower bound on accuracy** placed on the detector. One deficiency in our calculation of the Berg-Purcell limit is that we estimated but did not exactly calculate the M. Recent studies have proposed a method of exactly calculating the M. In doing so, these studies also suggested more elaborate mechanisms that ensure that each measurement is really independent. They include:

- K. Kaizu, W. de Ronde, J. Paijmans, K. Takahashi, F. Tostevin, and P.R. ten Wolde
  The Berg-Purcell limit revisited, *Biophysical Journal* (2014).
- W. Bialek and S. Setayeshgar
  Physical limits to biochemical sensing, *Proc. Natl. Acad. Sci. USA* (2005).

Despite the more elaborate calculations contained in these later studies, the bottom line is that the Berg-Purcell limit (Eq. 12) is still basically correct.

B. Applying the Berg-Purcell limit to the fruit fly embryo

In this section, we briefly summarize beautiful, physics-type high-precision experiments that Thomas Gregor and colleagues report in:

- T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek. Probing the limits to positional information, *Cell* (2007)

In this paper, Gregor et al. investigated how accurately nuclei inside the fruit fly embryo can measure the concentration of the morphogen called *Bicoid* that surrounds them. In the early embryo of the fruit fly, individual nuclei, not cells, are arranged next to each other in a near triangular lattice. The nuclei are not yet engulfed inside cell membranes - they are basically "cells" without membranes and other organelles. Simply put, the mother deposits the Bicoid mRNA at one end of the embryo. This end will form the head (i.e., *anterior*). Ribosomes that diffuse inside the embryo encounter and translate the mRNA into the Bicoid proteins. The Bicoid proteins diffuse from the head towards the other end of the embryo, which will form the tail (i.e., *posterior*). In so doing, a Bicoid-concentration gradient forms. Each nucleus in the embryo must "measure" the concentration of the Bicoid around itself. The concentration of the Bicoid that a nucleus measures is a function of the position of the nucleus relative to the embryo's head, where the Bicoid concentration would be the highest (since that is where the translation of Bicoid occurs - the mRNA is essentially localized at the embryo's head). Thus a nucleus can potentially "know" how far from the head it is by accurately measuring the Bicoid concentration. But due to the Berg-Purcell limit, there is a limit to how accurately each nucleus can measure the Bicoid concentration immediately surrounding it. Gregor et al. quantified how accurate is ?accurate enough? with calculations and tour de force, physics-type precision experiments on the fly embryo. For their experiments, Gregor et al. genetically engineered the fruit fly embryo in which each Bicoid protein is fused to a Green Fluorescent Protein (GFP) and built a high-resolution microscope called the **two-photon microscope** to measure the *number* of GFP molecules with a near integer-number precision (i.e., 1 GFP molecules, 2 GFP molecules, etc.). The latter was made possible by the fact that the two-photon microscope could resolve multiple Bicoid-GFP molecules that were very near each other as distinct molecules.

We now discuss the paper's main findings. Two hours after the fruit fly egg has been fertilized, the length L of the embryo is approximately 500 μm and the average distance Δx between the centers of two adjacent nuclei (which we treat as spheres) that are aligned along the straight line joining the head to tail (anterior-posterior) is approximately 8 μm (see Figure in the paper). Each nucleus has a diameter of approximately 5 μm. Based on these measurements, we have

$$\frac{\Delta x}{L} = \frac{8\ \mu\text{m}}{500\ \mu\text{m}} \approx 0.016 \tag{13}$$

Thus having two adjacent nuclei being able to tell that they are at two different locations along the anterior-posterior axis means that the nuclei must be able to resolve a difference in length by about 1.6% of the embryo's total length. Suppose that a nucleus is at position x where the Bicoid concentration is $c(x)$. Let $\Delta c + c(x)$ be the Bicoid concentration on its adjacent nucleus. Suppose that the Bicoid protein, after it is made at the anterior end of the embryo (i.e., at $x = 0$), diffuses freely inside the embryo towards the posterior end. Then solving the 1-dimensional diffusion equation with a source of molecules at the anterior yields

$$c(x) = c_0 e^{-x/\lambda} \tag{14}$$

where λ is the diffusion length. Gregor et. al. have measured the concentration of the Bicoid inside tens of fly embryos. In every embryo, they found that the concentration $c(x)$ indeed exponentially decayed as in Eq. 14 from the

anterior to posterior. Intriguingly, they found that there was very little variability in the concentration profile $c(x)$ from one embryo to another, despite coming from different mothers (in a later study, Gregor's lab has shown that all mother flies likely lay nearly the same, integer number of Bicoid mRNA at the anterior ends of their embryos!). Their measurements showed that the diffusion length λ is approximtaely 100 μm. For two adjacent nuclei, we have

$$\left|\frac{\Delta c}{c(x)}\right| = \left|\frac{1}{c(x)}\frac{dc}{dx}\right|\Delta x \tag{15a}$$

$$= \frac{\Delta x}{\lambda} \tag{15b}$$

$$= \frac{8\ \mu\text{m}}{100\ \mu\text{m}} \tag{15c}$$

$$\approx 0.10 \tag{15d}$$

According to Eq. 15d, the Bicoid concentration on a given nucleus is different by approximately 10% from the Bicoid conentration on its adjacent nucleus. Thus for two adjacent nuclei to "know" that they are apart by a distance of Δx, they must be able to distinguish 10% or less of a difference in their respective Bicoid concentrations.