BE/APh161: Physical Biology of the Cell
Homework 3
Due Date: Wednesday, January 27, 2021

"Champions aren't made in gyms. Champions are made from something they have deep inside them - a desire, a dream, a vision. They have to have the skill, and the will. But the will must be stronger than the skill." - Muhammad Ali

Extra Credit. Provide comments on chap. 4, "Thinking Big About Data" of the upcoming third edition of *Physical Biology of the Cell*. Note that this is an unfinished draft of the chapter. Figure placements are not necessarily correct and there are still a number of internal discussions amongst the author team about how to finish things off. We are especially interested in mistakes, flaws in logic, confusing figures, unclear discussions, etc., but are happy to entertain comments at all scales. This extra credit will constitute an additional 15% on your score on the homework.

1. Estimating the diffusion constant.

In this problem, we are going to use the observed trajectories of diffusing GFP molecules to estimate the diffusion coefficient.

(a) Conventional microscopy to observe individual fluorescent proteins won't work. In this part of the problem, we are going to work out why. During a traditional experiment, the microscope shutter is open during some time interval of order 10s to 100s of milliseconds. By assuming a diffusion constant of 10 μm^2/s, work out how far the fluorescent protein will move during the time that the shutter is open and compare that distance to the size of the cell itself and comment on how this limits our ability to measure the diffusion constant. Perform the estimate a second time, this time using the 0.3 ms exposure time shown in Figure 1(A).

(b) Using the trajectories shown in Figure 2 and our simple rule of thumb that $t_{diffusion} = L^2/D$ to estimate the diffusion constant for GFP. Explain your reasoning carefully.

2. The Standard Candle: Counting Proteins with Partitioning Statistics.

(a) Begin by reading the paper by Rosenfeld *et al.* entitled "Gene Regulation at the Single-Cell Level" (posted on the website with the homework) and write a one paragraph commentary on the paper with special reference to how they used the binomial partitioning as a way to count repressor proteins. What is the experiment they did and what were they trying to learn?

In the rest of the problem we work out for ourselves the ideas about binomial partitioning introduced in the Rosenfeld *et al.* paper in order to consider the concentration of mRNA or proteins as a function of time in dividing cells. In particular, the point of this problem is to work out the concentration of mRNA or protein given that we start with a single parental cell that has N copies of this mRNA or protein (in the experiments of Rosenfeld *et al.* this is a fluorescently-labeled transcription factor). In the Rosenfeld experiment, at some point while the culture is growing, the production of the protein is stopped by providing a chemical in the medium and then the number of copies per cell is reduced as a result of dilution as the cells divide.
Interestingly, this problem opens the door to one of the most important themes in physics, namely, that of fluctuations. In particular, as the cells divide from one generation to the next, each daughter does not really get $N/2$ copies of the protein since the dilution effect is a stochastic process. Rather the partitioning of the N proteins into daughter cells during division follows the binomial distribution. Analyzing these fluctuations can actually lead to a quantification of the number of copies of a protein in a cell.

(b) If we think of the N copies of the protein as being divided between the two daughters with N_1 going to daughter 1 and $N - N_1$ going to daughter 2, write the probability distribution $p(N_1, N)$. Next, work out the expected fluctuations in the partitioning process after each division by noting that the fluctuations can be written as $\sqrt{< (N_1 - N_2)^2 >}$, where N_1 and N_2 are the number of proteins that end up in daughter cells 1 and 2, respectively. Show that $\sqrt{< (N_1 - N_2)^2 >} = \sqrt{N}$. When I do this calculation, I find it convenient to write $N_2 = N - N_1$. Basically, this reduces the problem to having to calculate $\langle N_1 \rangle$ and $\langle N_1^2 \rangle$ since once you have those two quantities you can evaluate $\sqrt{< (N_1 - N_2)^2 >}$.

(c) Next, look at the Rosenfeld paper and explain how measuring fluorescence variations can be used to calibrate the exact number of copies of the fluorescent protein in a cell. Specifically, assume that the fluorescence intensity in each cell can be written as $I = \alpha N$, where α is an as-yet unknown calibration factor and N the number of proteins in the cell. Explain what this equation means and why you think it is justified. Derive an expression relating I_1, I_2 and I_{tot} using the result of part (b). Make a plot of $\sqrt{< (I_1 - I_2)^2 >}$ versus I_{tot} and explain how to get the calibration factor α from this plot.

(d) Now we are going to repeat the Rosenfeld experiment numerically in order to *fit* the calibration factor. Consider a fluorescent protein such that the calibration factor between the intensity and the number of fluorophores is 50, that is $I = 50N$. Generate intensity data by choosing $N_1 + N_2 = 10, 50, 100, 1000$ and 5000 and for each case, "partition" the proteins from the mother cell to the two daughters 100 times (i.e. as if you are looking at 100 mother cells divide for each choice of the protein copy number). Then, make a plot of the resulting $\sqrt{< (I_1 - I_2)^2 >}$ vs I_{tot} just as we did analytically in the previous problem. What I mean is that you need to make a plot of all of your simulation results. Then, do a fit to your "data" and see how well you recover the calibration factor that you actually put in by hand. Plot the fit on the same graph as all of the "data".

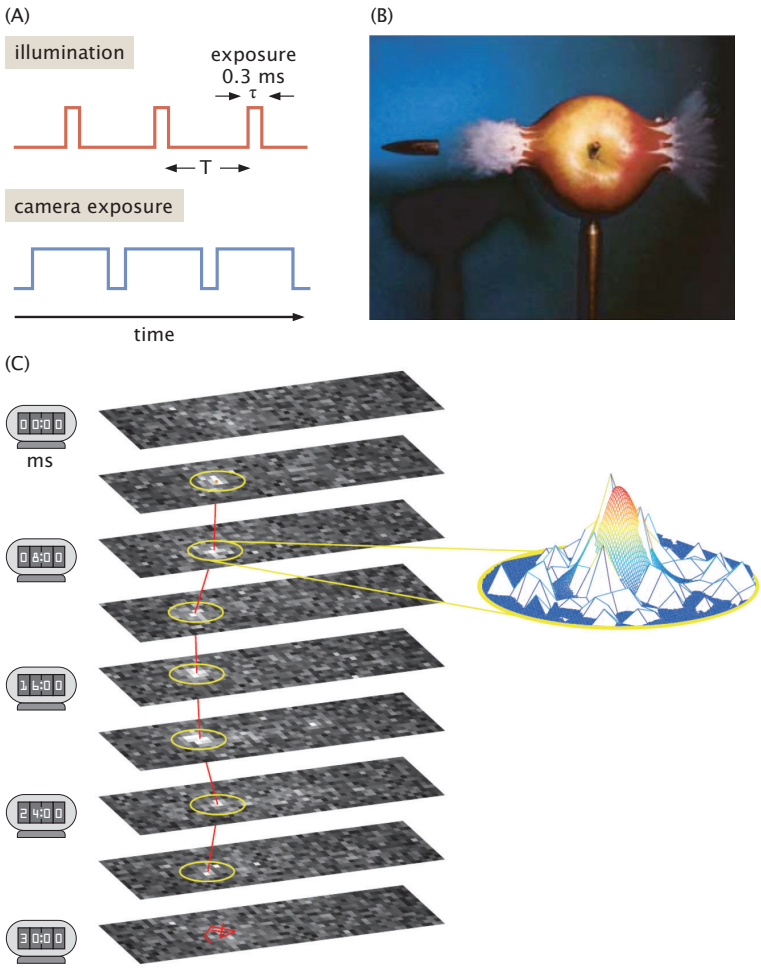

Figure 1: Stroboscopic illumination for high speed photography. (A) Comparison of the pulse of light and the camera exposure times. Brief illumination periods guarantee that the diffusing molecule doesn't move very far. (B) A classic photo from MIT legend Harold Edgerton who pioneered stroboscopic photography for science and fun. (C)

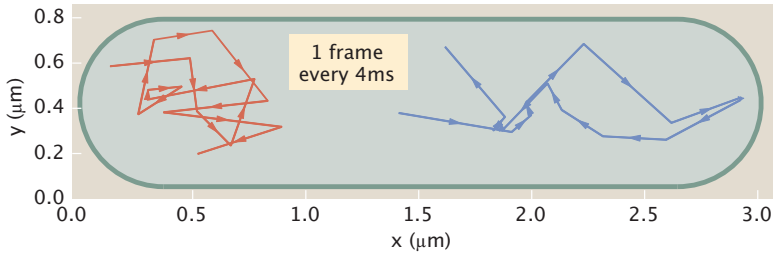

Figure 2: Time series showing positions of diffusing GFP molecules at different times. The red and blue traces correspond to different molecules. The lines are a guide to the eye.