II. NOISE ISN'T NEGLIGIBLE

The great poetic images of classical physics are those of determinism and clockwork. In a clock, not only the output but also the internal mechanisms are models of precision. Strikingly, life seems very different. Interactions between molecules involve energies of just a few times the thermal energy. Biological motors, including the molecular components of our muscles, move in elementary steps that are on the nanometer scale, driven forward by energies that are larger than the thermal energies of Brownian motion, but not much larger. Crucial signals inside cells often are carried by just a handful of molecules, and these molecules inevitably arrive randomly at their targets. Human perception can be limited by noise in the detector elements of our sensory systems, and individual elements in the brain, such as the synapses that pass signals from one neuron to the next, are surprisingly noisy. How do the obviously reliable functions of life emerge from under this cloud of noise? Are there principles at work that select, out of all possible mechanisms, the ones that maximize reliability and precision in the presence of noise?

In this Chapter, we will take a tour of various problems involving noise in biological systems. I should admit up front that this is a topic that always has fascinated me, and I firmly believe that there is something deep to be found in exploration of these issues. we will see the problems of noise in systems ranging from the behavior of individual molecules to our subjective, conscious experience of the world. In order to address these questions, we will need a fair bit of mathematical apparatus, rooted in the ideas of statistical physics. I hope that, armed with this apparatus, you will have a deeper view of many beautiful phenomena, and a deeper appreciation for the problems that organisms have to solve.

A. Molecular fluctuations and chemical reactions

In order to survive, living organisms must control the rates of many chemical reactions. Fundamentally, all reactions happen because of fluctuations. More strongly, chemical reactions are a non–perturbative consequence of molecular fluctuations. You all learned, perhaps even in high school, that the rates of chemical reactions obey the Arrhenius law, $k \propto e^{-E_{\rm act}/k_BT}$, where $E_{\rm act}$ is the activation energy. We also know that k_BT measures the mean square amplitude of fluctuations, for example in the velocities of atoms. Thus, chemical reaction rates are $\sim e^{-1/g}$, where g is the strength of the fluctuations. If we start by imagining a world in which there are no fluctuations, we can add them in piece by piece, but there is no way to get a chemical reaction rate as a perturbative series in g. Chemical reactions are so commonplace that we sometimes forget just how nontrivial they are from a theoretical point of view. Indeed, as I verify every year, few of the students in my course have ever seen an honest calculation that gives the Arrhenius law as a result, although they have all heard vague arguments about the Boltzmann probability of being on top of the barrier. So, our first order of business is to see how the Arrhenius law emerges, as an asymptotic result, for some real dynamical model. Only once we have this more solid understanding will we be ready to look at what might be special regarding the control of chemical reaction rates in biological systems.

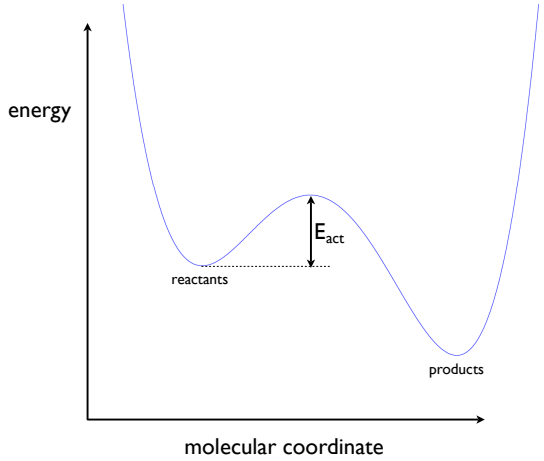

FIG. 32 The simplest model of a chemical reaction. Along some molecular coordinate x, the potential energy $V(x)$ has two minima separated by a barrier. The height of the barrier is the "activation energy" $E_{\rm act}$, which we expect will determine the rate of the reaction through the Arrhenius law, $k \propto e^{-E_{\rm act}/k_BT}$.

Let us consider the simplest case, shown in Fig 32. Here the molecules of interest are described by a single coordinate x, and the potential energy $V(x)$ as a function of this coordinate has two wells that we can identify as reactant and product structures. Let's assume that motions along this coordinate are overdamped, so inertia is negligible.[30] Since the molecule is surrounded by an environment at temperature T, we really want to describe Brownian motion in this potential. So, the equation of motion is[31]

$$\gamma \frac{dx}{dt} = -\frac{dV(x)}{dx} + \zeta(t), \tag{224}$$

[30] This really is just a simplifying assumption. We can also do everything in the case where inertia is significant, and none of the main results will be different. More precisely, we are going to go far enough to show that the Arrhenius law $k = Ae^{-E_{\rm act}/k_BT}$ is true, and that the activation energy $E_{\rm act}$ corresponds to our intuition. The neglect of inertia would only change the prefactor A, which is in any case much more difficult to calculate.

[31] For background on the description of random functions of time, see Appendix A.2.

where γ is the friction or drag coefficient, and the random or Langevin force $\zeta(t)$ reflects the random influences of all the other degrees of freedom in the system; to insure that the system eventually comes to equilibrium at temperature T we must have

$$\langle \zeta(t)\zeta(t')\rangle = 2\gamma k_B T \delta(t-t'). \tag{225}$$

The challenge is to see if we can extract from these dynamics some approximate result which corresponds to our intuition about chemical reactions, and in particular gives us the exponential dependence of the rate on the temperature.

[Perhaps should add some discussion of the "reaction coordinate" concept. On the other hand, one could say that we are just doing the simplest case, which is one dimensional, in which case there is no need for apologies, just generalization later. Advice welcome.]

When we solve Eq (224), what we get is the coordinate as a function of time. What features of this trajectory correspond to the reaction rate k? If there really are only two states in the sense of chemical kinetics, then trajectories should look like those in Fig 33. Specifically, we should see that trajectories spend most of their time in one potential well or the other, punctuated by rapid jumps between the wells. More precisely, there should be a clear separation of time scales between the dynamics within each well and the typical time between jumps. Further, if we look at the times spent in each well, between jumps, these times should be drawn from an exponential distribution, $P(t) = ke^{-kt}$, and then k is the rate constant for the chemical reaction leading out of that well into the other state.

Problem 38: What's the alternative? You should think a bit about what was just said. Suppose for example, that you don't know the potential and I just give you samples of the trajectory $x(t)$. What would it mean if the trajectories paused at some intermediate point between reactants and products? How would you interpret non–exponential distributions of the time spent in each well?

Problem 39: Numerical experiments on activation over a barrier. Perhaps before launching into the long calculation that follows, you should get a feeling for the problem by doing a small simulation. Consider a particle at position x moving in a potential $V(x) = V_0[1-(x/x_0)^2]^2$. Notice that this is double well, with minima at $x = \pm x_0$ and a barrier of height V_0 between these minima. Let's consider the overdamped limit of Brownian motion in this potential, as in Eq (224),

$$\gamma \frac{dx(t)}{dt} = \frac{4V_0}{x_0}\left(\frac{x}{x_0}\right)\left[1-\left(\frac{x}{x_0}\right)^2\right] + \zeta(t), \tag{226}$$

We want to simulate these dynamics. The simplest approach is the naive one, in which we use discrete time steps separated by Δt and we approximate

$$\frac{dx(t)}{dt} \to \frac{x(n+1)-x(n)}{\Delta t}. \tag{227}$$

(a.) To use this discretization we have to deal with the Langevin force. One (moderately) systematic approach is to integrate the Langevin equation over a small window of time Δt:

$$\gamma \int_t^{t+\Delta t} dt\, \frac{dx(t)}{dt} = -\int_t^{t+\Delta t} dt\, \frac{\partial V(x)}{\partial x} + \int_t^{t+\Delta t} dt\, \delta F(t) \tag{228}$$

$$\gamma\,[x(t+\Delta t) - x(t)] \approx -\Delta t \left.\frac{\partial V(x)}{\partial x}\right|_{x=x(t)} + z(t), \tag{229}$$

where

$$z(t) = \int_t^{t+\Delta t} dt\, \zeta F(t). \tag{230}$$

Using the correlation function of the Langevin force from Eq (225), compute the variance of $z(t)$. Show also that the values of z at different times—separated at least by one discrete step Δt—are uncorrelated.

(b.) Combine your results in [a] with the equations above to show that this simple discretization is equivalent to

$$y(n+1) = y(n) + \alpha E^\dagger \cdot y(n) \cdot [1-y^2(n)] + \sqrt{\frac{\alpha}{2}}\zeta(n), \tag{231}$$

where $y = x/x_0$, the parameter $\alpha = 4k_BT\Delta t/(\gamma x_0^2)$ should be small, $E^\dagger = V_0/(k_BT)$ is the normalized "activation energy" for escape over the barrier, and $\zeta(n)$ is a Gaussian random number with zero mean, unit variance, and no correlations among different time steps n.

(c.) Implement Eq (231), for example in MATLAB. Note that MATLAB has a command `randn` that generates Gaussian random numbers.[32] You might start with a small value of $E^\dagger$, and experiment to see how small you need to make α before the results start to make sense. What do you check to see if α is small enough?

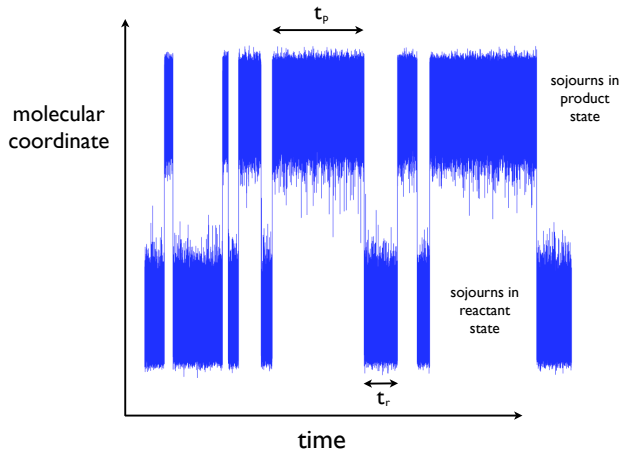

FIG. 33 Example of the trajectories we expect to see in solving the Langevin Eq (224). Long sojourns in the reactant or product state are interrupted by rapid jumps from one potential well to the other. If we look at the times t_r spent in the reactant state, these should come from a probability distribution $P_r(t_r) = k_+ e^{-k_+ t_r}$, where k_+ is the rate of the chemical reaction from reactants to products. Similarly we should have $P_p(t_p) = k_- e^{-k_- t_p}$, where k_- is the rate of the reverse reaction.

---

[32] More precisely, MATLAB *claims* that `randn` generates Gaussian random numbers that are independent. Maybe you should check this?

(d.) Explore what happens as you change the value of $E^\dagger$. For each value of $E^\dagger$, check that your simulation runs long enough so that the distribution of x actually is given by the Boltzmann distribution, $P(x) \propto \exp[-V(x)/k_BT]$. As $E^\dagger$ increases, can you see that there are isolated discrete events corresponding to the "chemical reaction" in which the system jumps from one well to the other? Use your simulation to estimate the rate of these jumps, and plot the rate as a function of the activation energy $E^\dagger$. Can you verify the Arrhenius law?

Problem 40: Effective potentials. We are discussing, for simplicity, a one dimensional problem. Suppose that there are really many dimensions, not just x but also y_1, y_2, $\cdots$, $y_N \equiv \{y_j\}$. Then we have, again in the overdamped limit,

$$\gamma \frac{dx}{dt} = -\frac{\partial V(x;\{y_j\})}{\partial x} + \zeta(t) \tag{232}$$

$$\gamma_i \frac{dy_i}{dt} = -\frac{\partial V(x;\{y_j\})}{\partial y_i} + \xi_i(t), \tag{233}$$

where, as usual

$$\langle \xi_i(t)\xi_j(t')\rangle = 2k_BT\gamma_i\delta_{ij}\delta(t-t'). \tag{234}$$

Imagine now that x moves much more slowly than all the $\{y_i\}$.

(a.) Verify that, from Eq (233), the stationary distribution of $\{y_i\}$ at fixed x is the Boltzmann distribution,

$$P(\{y_j\}|x) = \frac{1}{Z(x)} \exp\left[-\frac{V(x;\{y_j\})}{k_BT}\right]. \tag{235}$$

(b.) If x is slow compared with all the $\{y_j\}$, it is plausible that we should average the dynamics of x in Eq (232) over the stationary distribution $P(\{y_j\}|x)$. Show that this generates an equation in which x moves in an effective potential,

$$\gamma \frac{dx}{dt} = -\frac{\partial V_{\rm eff}(x)}{\partial x} + \zeta F(t), \tag{236}$$

and this effective potential is the free energy, $V_{\rm eff}(x) = -k_BT \ln Z(x)$.

(c.) Equations (232) and (233) still aren't completely general, since we have taken the mobility tensor to be diagonal, so that forces on coordinate y_i lead to velocities only along this direction. Does the more general case presents any new difficulties for the problem posed here?

This picture of trajectories that hover around one well and then jump to another should remind you of something you learned in quantum mechanics. In particular, if you take the path integral view of quantum mechanics, then tunneling in a double well potential is dominated by these sorts of trajectories. In fact, if Planck's constant is small, so that tunneling is rare, there is a semi–classical approximation to the path integral which reproduces the WKB approximation to Schrödinger's equation, and in this approximation the path integral is dominated by specific trajectories, which have come to be called "instantons." These instantons are precisely the jumps from one well to another, analogous to what we have drawn for the classical case in Fig 33.

There are three seemingly different but equivalent ways of doing quantum mechanics. Most elementary courses focus on Schrödinger's equation, which describes the amplitude for a particle to be at position x at time t. But you can also look at Heisenberg's equations of motion for the position (and momentum) operator, and finally one can use path integrals. How do these different approaches to quantum mechanics connect with the description of Brownian motion?

The Langevin equation is a bit like Heisenberg's equation for the position operator. It seems to give us something most closely related to the equations of motion in classical (noiseless) mechanics, but it requires some interpretation. In the case of the Langevin equation, because $\zeta(t)$ is random, when we solve for the trajectory $x(t)$ we get something different for every realization of ζ, so "solve" should be used carefully. More precisely what we get, for example, from one simulation of the Langevin equation is a sample drawn out of the distribution of trajectories.

When we pass from Heisenberg's equations of motion to the Schrödinger equation, we shift from trying to follow the time dependence of coordinates to trying to see the whole distribution of coordinates at each time, as encoded in the wave function. Similarly, we can pass from the Langevin equation to the diffusion equation, which governs the probability $P(x,t)$ that we will find the particle at position x at time t. It is useful to remember that the diffusion equation is an equation for the conservation of probability,

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x} J(x,t), \tag{237}$$

where $J(x)$ is the probability current.[33] Fick's law tells us that diffusion contributes a current that tends to reduce gradients in the concentration of particles, or equivalently gradients in the probability of finding one particle, so that

$$J_{\rm diff}(x,t) = -D\frac{\partial P(x,t)}{\partial x}. \tag{238}$$

But if there is some force $F(x) = -dV(x)/dx$ acting on the particle, it will move with an average velocity $v = F(x)/\gamma$, and hence there is a 'drift' current

$$J_{\rm drift}(x,t) = vP(x,t) = -\frac{1}{\gamma}\frac{dV(x)}{dx}P(x,t). \tag{239}$$

---

[33] A note about units. Often when discussing diffusion it is natural to think about the concentration of particles, which has units of particles per unit volume. The current of particles then has units of particles per are per time. What we are doing here is slightly different. First, we are talking about the probability of finding *one* particle at point x. Second, we are in one dimension, and so this probability distribution has units of 1/(length), not 1/(volume). Then the current has the units of a rate, 1/(time). Check that this make the units come out right in Eq's (**??**) and (238).

Putting these terms together, $J = J_{\rm diff} + J_{\rm drift}$, we have

$$\begin{aligned}\frac{\partial P(x,t)}{\partial t} &= -\frac{\partial}{\partial x}\left[-D\frac{\partial P(x,t)}{\partial x} - \frac{1}{\gamma}\frac{dV(x)}{dx}P(x,t)\right] \\ &= D\frac{\partial}{\partial x}\left[\frac{\partial P(x,t)}{\partial x} + \frac{1}{\gamma D}\frac{dV(x)}{dx}P(x,t)\right] \\ &= D\frac{\partial}{\partial x}\left[\frac{\partial P(x,t)}{\partial x} + \frac{1}{k_BT}\frac{dV(x)}{dx}P(x,t)\right],\end{aligned} \tag{240}$$

where in the last step we use the Einstein relation $D = k_BT/\gamma$. This way of writing the diffusion equation makes clear that the Boltzmann distribution $P \propto e^{-V(x)/k_BT}$ is an equilibrium ($\partial P/\partial t = 0$) solution.

We have said that, in looking at solutions of the Langevin equation, the signature of a "chemical reaction" with rate k is that the trajectories $x(t)$ will look like they do in Fig 33. What is the corresponding signature in the solutions of the diffusion equation? More precisely, even if we solve the diffusion equation to get the full $P(x,t)$ from some initial condition, what is it about this solution that corresponds to the rate constant k? In the same way that Schrödinger's equation is a linear equation for the wave function, the diffusion equation is a linear equation for the probability, which we can write as

$$\frac{\partial P(x,t)}{\partial t} = \hat{L}P(x,t). \tag{241}$$

All the dynamics are determined by the eigenvalues of the linear operator $\hat{L}$:

$$P(x,t) = \sum_{\rm n} a_{\rm n} e^{\lambda_{\rm n} t} u_{\rm n}(x) \tag{242}$$

$$\hat{L}u_{\rm n}(x) = \lambda_{\rm n} u_{\rm n}(x). \tag{243}$$

We know that one of the eigenvalues has to be zero, since if $P(x,t)$ is the Boltzmann distribution, $P \propto e^{-V(x)/k_BT}$, it won't change in time. Deviations from the Boltzmann distribution should decay in time, so all the nonzero eigenvalues should be negative.

Problem 41: Positive decay rates. We know that $P \propto \exp[-V(x)/k_BT]$ is a stationary solution of the diffusion Eq (240). To study the dynamics of how this equilibrium is approached, write

$$P(x,t) = \exp\left[-\frac{V(x)}{2k_BT}\right] Q(x,t). \tag{244}$$

(a.) Derive the equation governing $Q(x,t)$. Show that (by introducing factors of i in the right place) this can be written as

$$\frac{\partial Q(x,t)}{\partial t} = -A^\dagger A Q(x,t), \tag{245}$$

where the combination $A^\dagger A$ is a Hermitian operator. This gives an explicit version of Eq (241); explain why this implies that all the eigenvalues $\lambda_{\rm n} \le 0$.

(b.) For the case of the harmonic potential, $V(x) = \kappa x^2/2$, show that the operators $A^\dagger$ and A are the familiar creation and annihilation operators from the quantum harmonic oscillator. Use this mapping to find all the eigenvalues $\lambda_{\rm n}$. How do these relate to the time constant for exponential decay that you get from the noiseless dynamics [Eq (224) with $\zeta(t) = 0$]?

If we place the molecule in some configuration that is far from the local minima in each potential well, it will 'slide' relatively quickly into its relaxed configuration, and execute some Brownian motion around this sliding trajectory so that it samples the Boltzmann distribution within the well. This relaxation should be be described by some of the eigenvalues $\lambda_{\rm n}$, and these should be large and negative, corresponding to fast relaxation. In practice, we know that molecules in solution achieve this sort of 'vibrational relaxation' within nanoseconds if not picoseconds.

The statement that there is a chemical reaction at rate k means that, as a population of molecules comes to equilibrium, all the equilibration *within* the reactant or product states is fast, corresponding to time scales much shorter than $1/k$. On the much longer time scale $1/k$, there is equilibration between the reactant and product states. Thus, if we look at the whole spectrum of eigenvalues $\lambda_{\rm n}$ for the diffusion equation, one eigenvalue should be zero (as noted above), almost all the others should be very large and negative, while there should be one isolated eigenvalue that is small and negative—and this will be the reaction rate k, or more precisely the sum of the rates for the forward (reactants → products) and backward (products → reactants) reactions.

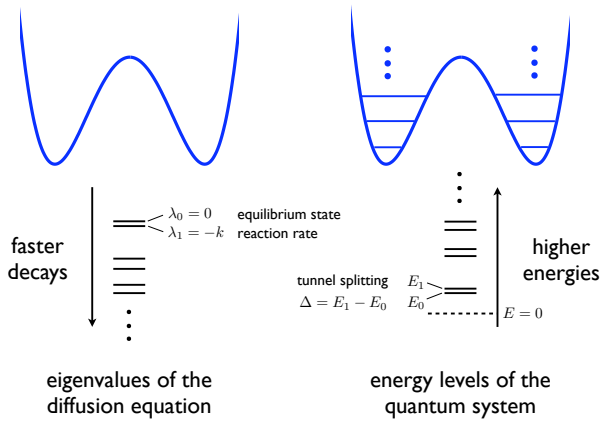

FIG. 34 Decay rates in diffusion compared with energy levels in quantum mechanics. In both cases there is a small splitting between the first two eigenvalues. For the diffusive case, this splitting is the rate of thermally activated hopping over the barrier—a chemical reaction. For the quantum case this splitting is the tunneling frequency between the two wells.

We arrive, then, at a picture of the eigenvalue spectrum in which there is a small splitting (between $\lambda_0 = 0$ and $\lambda_1 = -k$) relative to the next highest eigenvalue, as shown in Fig 34. This should remind you of what happens in quantum mechanical tunneling between two potential wells. The basic spacing of energy levels is set by the vibrational quanta within each well, but the states—and, in particular, the ground state—is split by a small amount corresponding to the frequency of tunneling between the two wells. It is the size of the barrier, or equivalently the smallness of $\hbar$, which makes this splitting small. In the diffusion problem, it is presumably the smallness of the temperature relative to the activation energy which enforces $\lambda_0 - \lambda_1 \ll \lambda_1 - \lambda_2$. We know how to solve Schrödinger's equation using the WKB approximation to extract the small tunneling amplitude, and so there should be a similar approximation to the diffusion equation that allows us to calculate the reaction rate.

The WKB approximation has a natural formulation in the path integral approach—in the limit $\hbar \to 0$, the path integral describing the amplitude for any quantum process is dominated by particular trajectories that are solutions of the classical equations of motion, although for classically forbidden processes (as with tunneling) these equations have to be continued to imaginary time. This idea of a dominant trajectory should be even clearer in the case of Brownian motion, since we won't have to deal with the continuation to imaginary time. To see how this works—and, finally, to derive the Arrhenius law—we need to construct the probability distribution functional for the trajectories $x(t)$ that solve the Langevin Eq (224).

The probability that we observe a trajectory $x(t)$ can be calculated by finding the random force $\zeta(t)$ which was needed to generate this trajectory, and then calculating the probability of this force. We know that the random forces come from a Gaussian distribution, and we know the correlation function [Eq (225)], so we have

$$P[\zeta(t)] \propto \exp\left[-\frac{1}{4\gamma k_B T}\int dt\, \zeta^2(t)\right]. \tag{246}$$

The Langevin equation, Eq (224), can be rewritten as

$$\zeta(t) = \gamma\frac{dx}{dt} + \frac{dV(x)}{dx}, \tag{247}$$

so it is tempting to say that the probability of observing the trajectory $x(t)$ is given by

$$P[x(t)] \sim \exp\left[-\frac{1}{4\gamma k_B T}\int dt\, \left(\gamma\frac{dx}{dt} + \frac{dV(x)}{dx}\right)^2\right], \tag{248}$$

and this is almost correct. To see what's missing, consider the simpler case where we just have one variable x [instead of a function $x(t)$] that obeys an equation

$$f(x) = y, \tag{249}$$

and y is random, drawn from a distribution $P_y(y)$. It is tempting to write

$$P_x(x) = P_y(y = f(x)), \tag{250}$$

but this can't be right—x and y can have different units, and hence P_x and P_y must have different units. As you have probably seen many times before, in this simple one dimensional example, the correct statement is that the probability mass within some small region dx must be equal to the mass found in the corresponding dy,

$$P_x(x)dx = P_y(y = f(x))dy \tag{251}$$

$$\Rightarrow P_x(x) = P_y(y = f(x))\left|\frac{dy}{dx}\right| \tag{252}$$

$$= P_y(y = f(x))\left|\frac{df(x)}{dx}\right|. \tag{253}$$

More generally, in order to equate probability distributions, we need a Jacobian for the transformation between variables. Thus, instead of Eq (248), we really want to write

$$P[x(t)] \propto \exp\left[-\frac{1}{4\gamma k_B T}\int dt\, \left(\gamma\frac{dx}{dt} + \frac{dV(x)}{dx}\right)^2\right]\mathcal{J}, \tag{254}$$

where $\mathcal{J}$ is the Jacobian of the transformation between $x(t)$ and $\delta F(t)$. Importantly, the Jacobian doesn't depend on temperature. In contrast, the exponential term that we have written out is $\sim e^{-1/T}$, so at low temperatures this will dominate. So, for this discussion, we won't worry about the Jacobian.

Problem 42: Jacobians. [Give a problem that walks through the derivation of the Jacobian, as in Zinn–Justin.]

To make use of Eq (254), it's useful to look more closely at the integral which appears in the exponential. Let's be careful to let time run from some initial time t_i up to some final time t_f:

$$\int_{t_i}^{t_f} dt \left(\gamma\frac{dx}{dt}+\frac{dV(x)}{dx}\right)^2 = \int_{t_i}^{t_f} dt \left[\left(\gamma\frac{dx}{dt}\right)^2 + 2\gamma\frac{dx}{dt}\frac{dV(x)}{dx} + \left(\frac{dV(x)}{dx}\right)^2\right] \quad (255)$$

$$= \int_{t_i}^{t_f} dt \left[\left(\gamma\frac{dx}{dt}\right)^2 + \left(\frac{dV(x)}{dx}\right)^2\right] + 2\gamma\int_{t_i}^{t_f} dt \frac{dx}{dt}\frac{dV(x)}{dx} \quad (256)$$

$$= \int_{t_i}^{t_f} dt \left[\left(\gamma\frac{dx}{dt}\right)^2 + \left(\frac{dV(x)}{dx}\right)^2\right] + 2\gamma\int_{t_i}^{t_f} dt \frac{dV(x)}{dt}, \quad (257)$$

$$= \int_{t_i}^{t_f} dt \left[\left(\gamma\frac{dx}{dt}\right)^2 + \left(\frac{dV(x)}{dx}\right)^2\right] + 2\gamma[V(x_f)-V(x_i)], \quad (258)$$

where in the last steps we recognize one term as a total derivative; as usual $x_i = x(t_i)$ is the initial position, and similarly $x_f = x(t_f)$ is the final position. Substituting, we can write the probability of a trajectory $x(t)$ as

$$P[x(t)] \propto \mathcal{J} e^{-S/k_BT}, \quad (259)$$

where the 'action' takes the form

$$S = \frac{V(x_f)-V(x_i)}{2} + \int_{t_i}^{t_f} dt \left[\frac{\gamma}{4}\left(\frac{dx}{dt}\right)^2 + \frac{1}{4\gamma}\left(\frac{dV(x)}{dx}\right)^2\right]. \quad (260)$$

This is a good time to remember that, for the simplest problems of classical mechanics, the action is

$$S_{cm} = \int_{t_i}^{t_f} dt \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 - \mathcal{U}(x(t))\right], \quad (261)$$

where m is the mass and $\mathcal{U}(x)$ is the potential energy. Except for a constant, the effective action for our problem is exactly that of a simple mechanics problem of a particle with mass m moving in a effective potential $\mathcal{U}(x)$,

$$m = \frac{\gamma}{2} \quad (262)$$

$$\mathcal{U}(x) = -\frac{1}{4\gamma}\left(\frac{dV(x)}{dx}\right)^2. \quad (263)$$

Figure 35 shows how this effective potential relates to the original double well.

At low temperatures, the distribution of trajectories will be dominated by those which minimize the action S. Clearly, one way to make the action minimal (zero, in fact) is to have the position be constant at one of the minima of the potential well. This describes a situation in which nothing happens. To have a chemical reaction, we need a trajectory that starts in the well corresponding to the reactants state, climbs up to the 'transition state' at the top of the barrier, and then slides down the other side. Let's start with the first part of this problem, finding a trajectory that climbs the barrier. The dominant trajectory of this form will be one that minimizes the action, and from Fig 35 we see that this is equivalent to finding the solution to an ordinary mechanics problem in which a particle starts on top of one hill, slides down and then gently comes to rest at the top of the next hill.

Problem 43: Zero energy? What we have just described are trajectories in the effective potential that have zero energy. There are, of course, trajectories that minimize the action but have nonzero energy. Why don't we consider these?

Taking the details of Fig 35 seriously, if we start at rest on top of one hill, this means that we start with zero energy. But energy is conserved along the trajectory, so that

$$\frac{m}{2}\left(\frac{dx}{dt}\right)^2 + \mathcal{U}(x(t)) = E = 0. \quad (264)$$

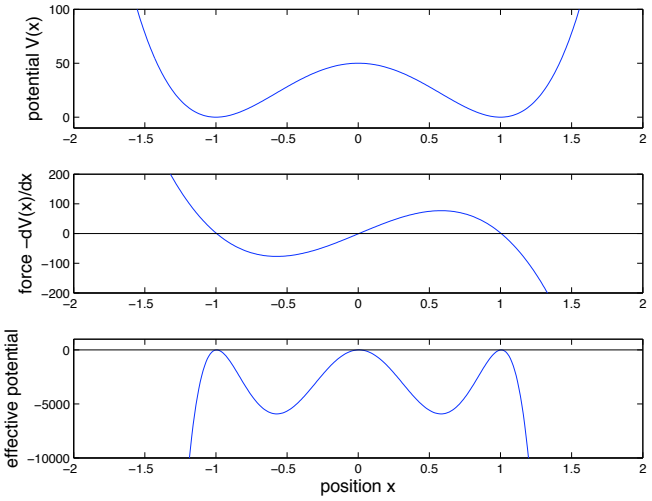

FIG. 35 Potentials and forces in a double well. Top panel show the true potential, middle panel the force, and bottom panel the effective potential that enters the probability distribution of trajectories. Notice that each extremum of the potential, both maxima and minima, becomes a maximum of the effective potential, and all these maxima are degenerate at $\mathcal{U} = 0$.

This means that

$$\mathcal{U}(x(t)) = -\frac{m}{2}\left(\frac{dx}{dt}\right)^2 \tag{265}$$

$$\frac{dx}{dt} = \pm\sqrt{-\frac{2}{m}\mathcal{U}(x(t))}; \tag{266}$$

we are interested in trajectories that move from left to right, so we should choose the upper sign, so that $dx/dt > 0$. But now we can substitute into the action,

$$\begin{aligned} S_{\rm cm} &= \int_{t_{\rm i}}^{t_{\rm f}} dt\, \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 - \mathcal{U}(x(t))\right] \\ &= \int_{t_{\rm i}}^{t_{\rm f}} dt\, \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 + \frac{m}{2}\left(\frac{dx}{dt}\right)^2\right] \qquad (267)\\ &= m\int_{t_{\rm i}}^{t_{\rm f}} dt\, \left(\frac{dx}{dt}\right)^2 \qquad (268)\\ &= m\int_{t_{\rm i}}^{t_{\rm f}} dt\, \frac{dx}{dt}\sqrt{-\frac{2}{m}\mathcal{U}(x(t))}, \qquad (269) \end{aligned}$$

so that finally we have

$$S_{\rm cm} = \int_{x_{\rm i}}^{x_{\rm f}} dx\, \sqrt{-2m\mathcal{U}(x)}, \tag{270}$$

where you should recognize the connection to the WKB formula for tunneling. In our case, the effective potential and mass are defined by Eq's (262) and (263), so that

$$-2m\mathcal{U}(x) = -2\frac{\gamma}{2}\left[-\frac{1}{4\gamma}\left(\frac{dV(x)}{dx}\right)^2\right] = \frac{1}{4}\left(\frac{dV(x)}{dx}\right)^2. \tag{271}$$

It is quite nice how the factors of γ cancel. Substituting into Eq (270), we find

$$\begin{aligned} S_{\rm cm} &= \frac{1}{2}\int_{x_{\rm i}}^{x_{\rm f}} dx\, \sqrt{\left(\frac{dV(x)}{dx}\right)^2} \qquad (272)\\ &= \pm\frac{1}{2}\int_{x_{\rm i}}^{x_{\rm f}} dx\, \frac{dV(x)}{dx} \qquad (273)\\ &= \frac{1}{2}\Big|V(x_{\rm f}) - V(x_{\rm i})\Big|, \qquad (274) \end{aligned}$$

where we choose the sign in taking the root so that the action comes out positive, as it must from Eq (268).

Problem 44: Extracting the dominant paths. We have seen that, in the low temperature limit, the reaction is dominated by trajectories that lead from one well to the other and minimize the action. Look through your simulation results from Problem 28, and collect as many examples as you can of the 'jumping' trajectories. How do these examples compare with the theoretical prediction that comes from minimizing the action? Can you align the sample trajectories well enough to compute an average that might be more directly comparable to the theory?

To finish the calculation, we need to put some of these pieces together. The action that determines the probability of a trajectory is, from Eq (260),

$$\begin{aligned} S &= \frac{V(x_{\rm f}) - V(x_{\rm i})}{2} + \int_{t_{\rm i}}^{t_{\rm f}} dt\, \left[\frac{\gamma}{4}\left(\frac{dx}{dt}\right)^2 + \frac{1}{4\gamma}\left(\frac{dV(x)}{dx}\right)^2\right] \\ &= \frac{V(x_{\rm f}) - V(x_{\rm i})}{2} + S_{\rm cm} \qquad (275)\\ &= \frac{V(x_{\rm f}) - V(x_{\rm i})}{2} + \frac{1}{2}\Big|V(x_{\rm f}) - V(x_{\rm i})\Big|. \qquad (276) \end{aligned}$$

This is a remarkably simple result. If we are looking at a trajectory that climbs from the bottom of a potential well to the top of the barrier, we have $V(x_{\rm f}) > V(x_{\rm i})$ and hence the action is

$$S_{\rm climb} = V(x_{\rm f}) - V(x_{\rm i}) = E_{\rm act}, \tag{277}$$

which is the "activation energy" for going over the barrier. On the other hand, if we look at a trajectory that slides down from the barrier into the other well, we have $V(x_{\rm f}) < V(x_{\rm i})$ and hence

$$S_{\rm slide} = 0. \tag{278}$$

So, what we have shown is that paths which take us from reactants to products, climbing the barrier and sliding down the other side, have a minimal action $S_{\rm react} = S_{\rm climb} + S_{\rm slide} = E_{\rm act}$. Thus, the probability of seeing such a trajectory is

$$P[x_{\rm react}(t)] \propto \mathcal{J}e^{-S_{\rm react}/k_BT} \sim e^{-E_{\rm act}/k_BT}, \tag{279}$$

and this is the essence of the Arrhenius law (at last).

One could legitimately complain that we haven't really solved our problem. All we have done is to show that, in some window of time, trajectories that jump from reactants to products are suppressed in probability by a factor $e^{-E_{\rm act}/k_BT}$. This is the basic idea of the Arrhenius law, but we haven't actually calculated a rate constant. In truth, this last step requires rather more technical apparatus, in the same way that getting the tunneling rate in the WKB approximation is harder than getting the exponential suppression, so I will leave it aside for now.

So far, we have given a fairly general discussion, and perhaps it's not obvious whether there is anything special about how these ideas will play out in the case of biological molecules. If we try to draw the picture in Fig 32, we usually associate the "reaction coordinate," that is the molecular coordinate along which we see the double well potential, with the motions that are involved in the chemical events themselves. Thus, if we are looking at the transfer of a hydrogen atom, breaking one bond and forming another, we might think that the relevant molecular coordinate is given by the position of the hydrogen atom itself.

Biological molecules—such as the proteins which act as enzymes, catalyzing specific chemical reactions of importance in the cell—are large, and hence flexible. Certainly they can change reaction rates by holding the reactants in place. But because of their flexibility, there is also the possibility that, as they flex, the effective barrier for the reaction changes. In this case, the dominant path for the reaction might be for the protein to fluctuate into a favorable configuration, and then for the more local coordinates (e.g., the position of the hydrogen atom) to make their jump. In this way, the observed activation energy comes to have two components, the usual one that we measure along the reaction coordinate, which presumably is reduced by waiting for the protein to arrange itself properly, and then the energy of distorting the protein itself.

To be a little more formal, imagine that for every configuration Q of the protein, there is a different activation energy for the reaction, $E_{\rm act}(Q)$. Of course there must also be some (free) energy of the protein once it is in the structure described by Q, and this determines the probability distribution $P(Q)$. Then if the fluctuations in Q are fast, we expect to see an average rate constant

$$k = A\int dQ\, P(Q) \exp\left[-\frac{E_{\rm act}(Q)}{k_BT}\right]. \qquad (280)$$

If we fix Q at its equilibrium position, we could find that $E_{\rm act}(Q = Q_{\rm eq})$ is large, which might make us think that the reaction will be slow. But by sampling non–equilibrium configurations, the protein can speed up the reaction.

Obviously this general picture depends on many details, but before proceeding one could ask if there is any evidence for such coupling of protein structural fluctuations to the modulations of chemical reaction rates. I think the strongest evidence is from the mid 1970s, in a beautiful series of experiments by Austin and colleagues. The idea is very simple. Suppose that we really do have the activation energy varying with the configuration of the protein. If we could stop the motion of the protein, then each molecule would be stuck with a different activation energy and hence a different reaction rate. Then, instead of seeing an average rate, each molecule reacts at its own rate, and if we count the total number of molecules that have not yet reacted we should see

$$N(t) = \int dQ\, P(Q) \exp\left[-Ae^{-E_{\rm act}(Q)/k_BT}t\right], \qquad (281)$$

which definitely is not an exponential decay. In fact if the fluctuations in Q generate very large variations in the activation energy, then this is very far from being an exponential decay.

Problem 45: Power law decays. Suppose that the effect of the fluctuations in Q is to generate a distribution of activation energies

$$P(E_{\rm act}) = \frac{1}{n!E_0}(E/E_0)^n e^{-E/E_0}. \qquad (282)$$

Then we should have

$$N(t) = \int_0^\infty \frac{dE}{n!E_0}(E/E_0)^n e^{-E/E_0} \exp\left[-Ae^{-E/k_BT}t\right]. \qquad (283)$$

(a.) Show that, at large t, there is a saddle point approximation to this integral, and that this predicts a decay $N(t) \sim t^{-\alpha}$. What determines the power α? Are there corrections to this formula?

(b.) Calculate the average rate constant, as in Eq (280),

$$\bar{k} = A\int_0^\infty \frac{dE}{n!E_0}(E/E_0)^n e^{-E/E_0} \exp\left[-\frac{E}{k_BT}\right]. \qquad (284)$$

Does this mean rate obey the Arrhenius law? How large are the deviations? Is there a limit in which the Arrhenius law is recovered?

Problem 46: A more sophisticated model. The discussion above concerns either the limit in which fluctuations are very fast, so we see an average rate constant, or very slow, so that we see a distribution of rate constants. It would be nice to have a simple model that interpolates between these limits. [give a problem that goes through the essence of the Agmon & Hopfield papers ...]

So, we have the dramatic prediction that if we would freeze the motion of the protein, we'd see something very far from the usual exponential decays. In order to test this we need the right model system. In particular, if we are literally going to freeze things, then molecules can't diffuse relative to one another, and most what we usually think of as chemistry will stop. We need an example of a reaction that happens among molecules that are already "together" and ready to react. If things are frozen, then

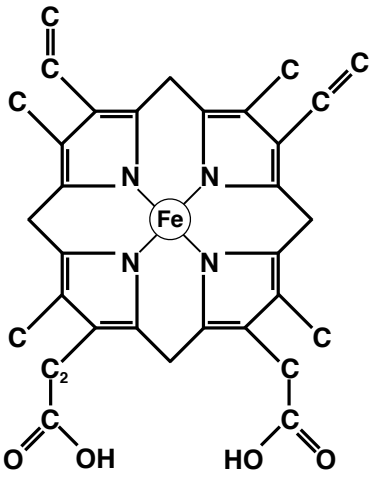

FIG. 36 The heme group at the center of myoglobin, hemoglobin, and other heme proteins. Recall the convention (Fig 15) that carbon atoms are at unmarked nodes of the skeleton, and hydrogen atoms which complete the four bonds needed for each carbon are not shown. The iron atom at the center is also coordinated from below by a nitrogen from the protein, and oxygen or carbon monoxide can bind to the iron from above the plane. The large conjugated structure of the heme group endows the molecule with a strong absorption cross section in the visible and ultraviolet range of the spectrum. Because the electronic states of the heme mix with the d orbitals of the iron, the absorption spectrum shifts upon oxygen binding.

the usual trick of suddenly mixing the reactants together to start the reaction also isn't going to work.

In many organisms, including us, oxygen is essential for a wide variety of processes. We take in oxygen by breathing, and need to distribute it to all of our tissues. The way we do this is to have specific proteins to which oxygen binds, and then the proteins are transported, starting in the blood. The major such oxygen transport protein in our blood is called hemoglobin, which is described in more detail in Appendix A.4 because it provides the classic example of cooperativity in protein function. Hemoglobin has four protein subunits, each of which can bind a single oxygen molecule. In our muscles we find a simpler protein, with just one subunit, called myoglobin. Myoglobin, hemoglobin, and the cytochromes that we will discuss below all are members of the "heme protein" family, which are defined by the fact that they bind a rather large planar organic molecule called heme, with an iron atom at its center, as shown in Fig 36. This iron is held in the plane by nitrogens from the heme and from below by a nitrogen from the protein. Oxygen can bind to the iron from above the plane.

The iron atom, and hence the oxygen binding site is buried deep inside the protein, as shown in Fig 37. This is interesting in part because it tells us that the full process of binding and unbinding must involve some motion or "breathing" of the protein structure. Further, once oxygen binds, if we freeze the protein it will be trapped, unable to escape. The conjugated electronic structure of the heme generates a strong optical absorption band, and because the electronic states of the heme mix with the orbitals of the iron, the absorption shifts when oxygen binds to the iron. Further, when a photon is absorbed by myoglobin with oxygen bound, there is some probability that the energy of the absorbed photon will be channeled into breaking the bond between the iron and the oxygen. Thus, if we let oxygen bind to myoglobin and then freeze the solution, we can knock the oxygen off the iron atom with a flash of light, and then we can watch the oxygen rebind after rattling around in the "pocket" formed by the protein.

In principle, motion of the oxygen molecule from the pocket to the iron atom needn't be coupled to motions of the protein. But if this coupling does occur, we expect, from the discussion above, that the kinetics of the rebinding after a light flash will deviate strongly from an exponential decay. We can follow the kinetics by looking at the absorption spectrum, and this is what is shown in Fig **??** for both oxygen and carbon monoxide binding to myoglobin. We see that once the solution is truly frozen solid (below $\sim 160\,\mathrm{K}$ in the glycerol–water mixtures used for these experiments), the fraction of molecules that have not reacted decays more nearly as a power law than an exponential. This suggests that we have frozen in a very

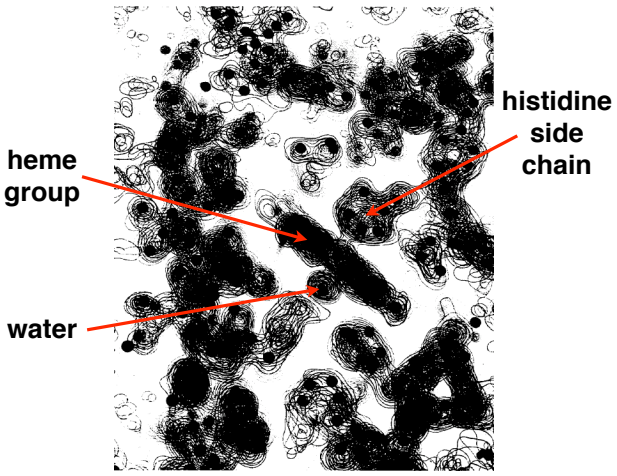

FIG. 37 A slice through the electron density map of the myoglogin molecule, as inferred from X–ray diffraction data (Kendrew 1964). This map is made from data at 1.4 Å resolution. In the center we see the heme group edge on. The histidine side chain from the protein coordinates the iron atom from below the plane of the heme, and in the crystals used in these experiments a water molecule binds to the iron atom in the position that would be taken by oxygen. Note that there is not much empty space in the structure, so that the protein actually has to "breathe" in order for oxygen to have access to the iron, or to escape once bound.

broad distribution of rate constants.

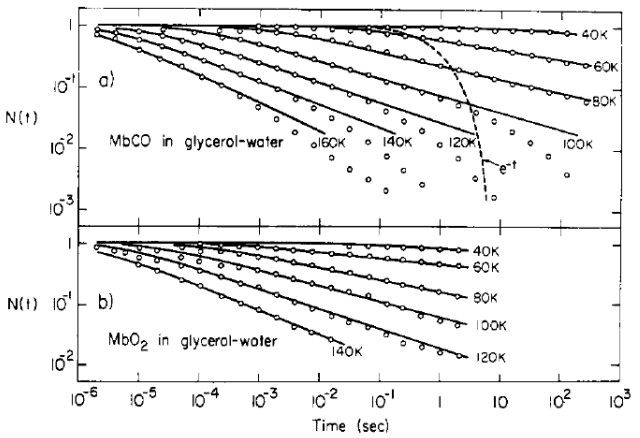

FIG. 38 Rebinding of oxygen and carbon monoxide to myoglobin at low temperatures, following a flash of light to break the bond (Austin et al 1975). Circle are data points, obtained by monitoring the absorption spectrum. Note that this is a logarithmic plot on *both* axes, so that we see an enormous range of times. Lines are fits to the phenomenological power law decay $N(t) = 1/[1 + (t/t_0)]^n$. The dashed line shows, for contrast, an exponential decay, $N(t) = e^{-kt}$, with $k = 1\,\mathrm{s}^{-1}$.

So far our discussion of chemical reactions has treated motion along the reaction coordinate as being completely classical. Is it possible that quantum effects could be relevant? Notice in Fig 38 that as we lower the temperature, the kinetics remain consistently non–exponential, but the typical time scale (e.g., the time required for the reach to reach 90% completion) is slowing down. If we keep lowering the temperature, eventually this slowing stops, and we see temperature independent kinetics. Almost certainly this arises because the reaction proceed by quantum mechanical tunneling through the effective barrier rather than by thermal activation over the barrier. The observation of quantum mechanical effects in a biological system always triggers excitement, although this is tempered somewhat by the fact that, in this case, to see tunneling one has to go to very low temperatures (below 10 K) indeed. In fact, well before the work on myoglobin, there had been observations of temperature independent kinetics in the photon–triggered electron transfer reactions in photosynthesis. Although our immediate experience of photosynthesis involved plants, many of the key experiments on the dynamics of electron transfer were done in photosynthetic bacteria.

The basic business of photosynthesis is carried out by the reaction center, a huge complex of proteins that holds a collection of medium sized, organic molecules—chlorophylls, pheophytins (chlorophylls without the magnesium), and quinones. [Need some schematics of these molecules, plus the reaction center structure.] Two of the chlorophylls are held in a special pair (P), and the electronic states of these two molecules are strongly mixed. If one purifies the reaction center away from all the accessory structures, the photochemistry is triggered when the special pair absorbs a photon.

From the excited state of the special pair, an electron hops to states localized on the pheophytin (I) and then the quinone (Q), as shown in Fig 39. Because P and Q are held, by the protein scaffold, on opposite sides of a membrane, the net effect is to transfer charge across the membrane, capturing the energy of the absorbed photon. Quinones [point back to the structure!] exist in multiple protonation states, so that the electron transfer can couple to proton transfer, and in this way the reaction center serves to drive protons across the membrane. The difference in electrochemical potential for protons provides a universal energy source that is used by other transmembrane protein, for example to synthesize ATP, which all cells use in powering other processes (including movement). In more complex organisms, including green plants, there are two kinds of reaction centers, one of which couples photon–driven electron transfer to the splitting of water to make all the oxygen in our atmosphere.

To complete the cycle and "reset" the reaction center for the arrival of the next photon, the hole on the special pair needs to be filled in, and this happens by electron transfer from another protein, cytochrome c, which can also diffuse away from the membrane and interact with the rest of the cell's chemistry. It is this reaction that

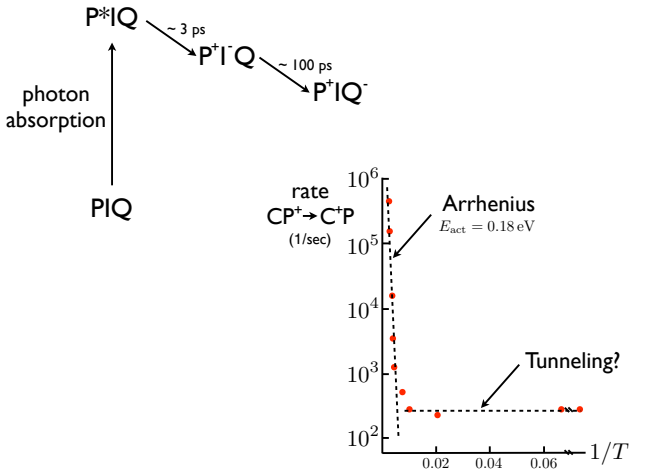

FIG. 39 Schematic of the electron transfer reactions in the reaction center of photosynthetic bacteria. The "pigment molecule" P absorbs a photon, and transfers an electron from the excited state to an intermediate acceptor I, which then passes the electron to a quinone molecule Q; there is a second quinone, not shown. The hole on P is filled in by electron transfer from another protein, cytochrome c, and the kinetics of the reaction $CP^+ \to C^+P$ provided the first evidence for quantum tunneling in a biological system, as shown (redrawn from DeVault & Chance 1966).

provided the first evidence for tunneling in a biological system. If the cytochrome is absent, as in purified reaction centers, one can observe the recombination reaction $P^+Q^- \rightarrow PQ$, which also has an anomalous temperature dependence, as discussed below. To connect with the discussion of myoglobin, this recombination reaction also exhibits non–exponential kinetics under some conditions, suggesting that it is possible to freeze some of the fluctuations in structure that normally provide rapid modulations of the reaction rate.

The key to experiments on the kinetics of photosynthetic electron transfer is that all of the molecules involved change their absorption spectra significantly when they gain or lose an electron; not coincidentally, these spectra overlap with the spectrum of the solar radiation, and are concentrated in a range of wavelengths surrounding the 'visible,' from the near infrared to the near ultraviolet. We can trigger the reactions with a pulsed laser tuned to the absorption band of P, and we can then monitor different spectral features that track the different components. This started in the 1950s and 60s with time resolution in the microsecond range, and evolved—with successive revolutions in the techniques for generating short laser pulses—down to picoseconds and femtoseconds; this development parallels the exploration of the visual pigments described in Section I.B.

One key point about the photosynthetic reaction center is that all the electron transfer processes work even when the system is frozen, which tells us that there is no need for the different components to diffuse in order to find one another—all of the donor and acceptor sites are held in place by the protein scaffolding. This allows for investigation of the electron transfer reactions over a wide range of temperatures, and this was done to dramatic effect by DeVault and Chance in the mid 1960s, with the result shown in Fig 39. Near room temperature, the electron transfer from cyctochrome c back to the special pair exhibits a normal, Arrhenius temperature dependence with an activation energy $E_{\rm act} \sim 0.18\,{\rm eV}$. Importantly, the temperature dependence is continuous as the system is cooled through the solvent's freezing point. But somewhere around $T \sim 100\,{\rm K}$, the temperature dependence stops, and the reaction rate remains the same down to liquid helium temperatures ($T \sim 4\,{\rm K}$). This strongly suggests that the reaction proceeds by tunneling at low temperatures.

Problem 47: A wrong model. If a reaction proceeds by activation over a barrier of height E, the rate is $k \propto \exp(-E/k_BT)$, If it proceeds by tunneling through the barrier, we expect $k \propto \exp(-2\sqrt{2mE}\ell/\hbar)$, where ℓ is the width of the barrier and m is the effective mass of the tunneling particle. For the DeVault–Chance reaction, there is a direct measurement of the activation energy $E \sim 0.18\,{\rm eV}$. If you imagine that it is the electron which has to go over or through this barrier, what value of ℓ is needed to explain that the crossover from Arrhenius behavior to temperature independence occurs near $T \sim 100\,{\rm K}$? Does this result make any sense?

After roughly a decade of confusion (including discussions of the model in the previous problem), a clearer understanding of tunneling in electron transfer emerged in the mid to late 1970s.[34] The basic idea is schematized in Fig 40. We have an electron donor D and an acceptor A; the reaction is $DA \rightarrow D^+A^-$. The states DA and D^+A^- are different *electronic* states of the system. From the Born–Oppenheimer approximation, we know that when a molecule shifts to a new electronic state, the nuclei move on a new potential surface. We usually describe these nuclear or atomic motions as molecular vibrations, so we'll refer to the relevant coordinates as vibrational coordinates. The simplest scheme, as in Fig 40, is one in which the vibrations are approximately harmonic. Then when we change electronic states, we can imagine changes in the structure of the normal modes, changes in the frequencies of these modes, and shifts in the equilibrium positions along the modes; barring symmetries, the last effect should be the leading one, and certainly it is the simplest.

In the state DA, an electron is localized on the donor. In the state D^+A^-, this electron is localized on the acceptor. If the donor and acceptor sites are far apart, as is often the case in large, biological molecules, then the

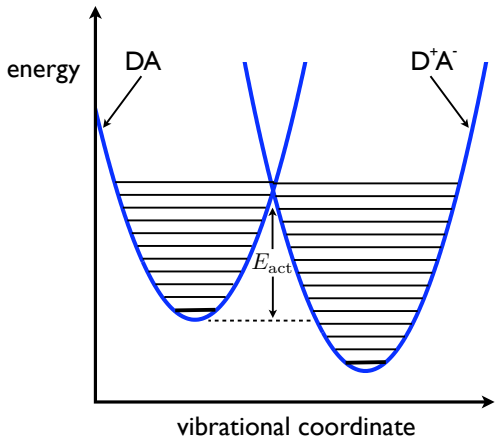

FIG. 40 Electron transfer is coupled to vibrational motion. ...

---

[34] The relevant physics here is essentially the same as in the discussion of absorption spectra in large molecules. See Chapter One and the Appendix on electronic transitions in large molecules; give pointer to arXiv version.

wave functions of the electrons in these localized state will overlap only deep in their tails; any matrix element that connects these two states then must be very small. But if we want to have a transition between two states that are connected by only a small matrix element, then by Fermi's golden rule we need these states to be of the same energy. As shown in Fig 40, this happens only at special points, where the two potential energy surfaces for vibrational motion cross. The rate of the reaction should then be proportional to the probability of finding the system at this crossing point. The key point, then, is that at high temperatures this probability is controlled by the thermal fluctuations in the vibrational coordinates, while at low temperatures the system can still reach the crossing point, but now the fluctuations are dominated by quantum zero–point motion. If the activation energy—the energy required to distort the molecule from its equilibrium structure in state DA to the crossing point—is large compared to the relevant vibrational quanta, then a zero–point fluctuation that carries the system to the crossing point necessarily involves sampling the tails of the ground state wavefunction, and this means that the system moves into a region that would be forbidden to a classical particle, even granting that it has the zero–point energy to work with. Thus, at low temperatures, the reaction is controlled by tunneling of the vibrational degrees of freedom, while at high temperatures these degrees of freedom move classically over the barrier.

To make all this a bit more precise, let's write the Hamiltonian corresponding to Fig 40. We have two electronic states, which we can take as the up and down states of a spin one–half. There is an energy difference between these states, which we'll call ϵ, and a weak matrix element Δ that mixes these states. There is a vibrational coordinate Q, and this coordinate moves in a potential that depends on the electronic state. Thus we have

$$\mathbf{H} = \frac{\epsilon}{2}\sigma_z + \Delta\sigma_x + \frac{1}{2}\dot{Q}^2 + \frac{1+\sigma_z}{2}V_\uparrow(Q) + \frac{1-\sigma_z}{2}V_\downarrow(Q), \tag{285}$$

If we think semi–classically, then the vibrational coordinates move hardly at all during the electronic transition, and so from the golden rule we should have the reaction rate

$$k \sim \frac{1}{\hbar}\Delta^2\left\langle \delta(E_\uparrow - E_\downarrow)\right\rangle = \frac{1}{\hbar}\Delta^2\left\langle \delta\left[\epsilon + V_\uparrow(Q) - V_\downarrow(Q)\right]\right\rangle, \tag{286}$$

where we have to average over the fluctuations of Q in the initial state DA. In the simplest case, where the potential surfaces are harmonic, differing only in their equilibrium positions,

$$V_\uparrow(Q) = \frac{\kappa}{2}Q^2 \tag{287}$$

$$V_\downarrow(Q) = \frac{\kappa}{2}(Q-Q_0)^2, \tag{288}$$

and hence $V_\uparrow(Q) - V_\downarrow(Q) = \kappa(Q_0 Q - Q_0^2/2)$, so that

$$k \sim \frac{1}{\hbar}\Delta^2\left\langle \delta\left(\epsilon - \frac{\kappa}{2}Q_0^2 + \kappa Q_0 Q\right)\right\rangle \tag{289}$$

$$= \frac{\Delta^2}{\hbar\kappa Q_0}P\left(Q = \frac{Q_0}{2} - \frac{\epsilon}{\kappa Q_0}\right). \tag{290}$$

If we have a particle moving in a harmonic potential with frequency ω, then in thermal equilibrium the distribution of Q is Gaussian. The variance is $\langle(\delta Q)^2\rangle = k_B T_{\rm eff}/\kappa$, where

$$k_B T_{\rm eff} = \hbar\omega\left[\frac{1}{2} + \frac{1}{e^{\hbar\omega/k_B T} - 1}\right]; \tag{291}$$

notice that as $T \to 0$, $k_B T_{\rm eff}$ approaches the zero–point energy $\hbar\omega/2$. Putting all the terms together, we find

$$k \sim \frac{\Delta^2}{\hbar\sqrt{4\pi\lambda k_B T_{\rm eff}}}\exp\left[-\frac{(\epsilon-\lambda)^2}{4\lambda k_B T_{\rm eff}}\right], \tag{292}$$

where $\lambda = \kappa Q_0^2/2$ is the "reorganization energy" that would be required to distort the molecule from its equilibrium configuration in DA into the equilibrium configuration appropriate to D^+A^- if we didn't actually transfer the electron.

In Figure 41 we see the predicted dependence of the electron transfer rate on temperature in a parameter regime chosen to match the DeVault–Chance reaction. In order to have the transition between Arrhenius and tunneling behavior at the right temperature, we need a vibrational frequency $\omega/2\pi \sim 200\,{\rm cm}^{-1}$.[35] If we look at the Raman spectra of cytochrome c or related molecules,

[35] Molecular vibrations contribute to the absorption of radiation in the infrared, and it is conventional to measure frequency in "wavenumbers" or inverse cm. To convert to the more usual Hz, just multiply by the speed of light, $3\times10^{10}\,$cm/s. Note that this is a convention about units, and *not* a reference to the inverse wavelength in the medium used for the experiment, so

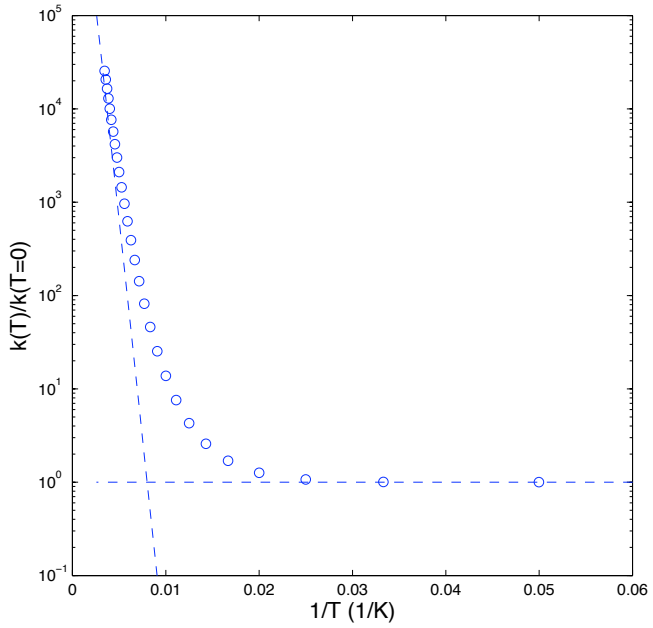

FIG. 41 Temperature dependence of the electron transfer rate, from Eq (292). Parameters are chosen, as described in the text, to match the behavior of the DeVault–Chance reaction in Fig 39. Circles are values of the rate computed at 20 K intervals, and dashed lines indicate the asymptotic behavior at high (activated) and low (tunneling) temperatures.

there is a vibrational mode near this frequency that corresponds to motions of the iron atom perpendicular to the plane of the heme group [obviously need a structural schematic!]. This makes sense, since when we add or subtract an electron from the molecule, this charge is shared between the iron and the heme, and on average the iron is displaced relative to the heme when the molecule changes its oxidation state. The energy difference between reactants and products can be measured directly by separate electrochemical experiments, and then to get the activation energy right we must have $\lambda \sim 0.14\,\text{eV}$. If the relevant vibrational mode really is (mostly) the motion of the iron relative to the rest of the protein, then we know the mass associated the mode and hence the stiffness $\kappa = m\omega^2$, so we can determine $Q_0 \sim 0.2\,\text{Å}$, and this is consistent with the displacements found upon comparing the oxidized and reduced structures of cytochrome c. So, this account of vibrational motion as controlling the temperature dependence of the reaction rate seems to make sense in light of everything else we know about these molecules, although admittedly it is a rough comparison. [Say something about the charge transfer band as a direct test?]

Problem 48: Getting numbers out. Convince yourself that the numbers in the preceding paragraph make sense. In particular, extract the estimate $Q_0 \sim 0.2\,\text{Å}$ for the motion of the iron atom relative to the protein.

There are many loose ends here. To begin, we have given a description in terms of one vibrational mode, but we have found an expression for the reaction rate that shows no sign of resonances when the energy difference ϵ between reactants and products in an integer multiple of the vibrational quantum $\hbar\omega$. Presumably the solution to the problem is the same as in our discussion of the absorption spectra of rhodopsin: individual modes are damped, so that resonances are broadened, and there are many modes, so the broadened resonances overlap and smear into a continuum.

The second problem concerns the significance of all this for biological function. It's very impressive to see quantum tunneling in a biological molecule, but our excitement should be tempered by the fact that we see this only at temperatures below 100 K, far out of the range where life actually happens. Measurements on the (much faster) initial steps of electron transfer, however, show that approximately temperature independent reaction rates persist up to room temperature. Indeed if we look closely at the rates of $P^*I \to P^+I^-$ and $I^-Q \to IQ^-$, we see a slightly inverse temperature dependence, with the rate slowing by a factor of two or three as we increase the temperature from 4 to 300 K [should have a figure for this!]. In fact the theory as we have sketched it provides a possible explanation for this: if we tune the energy difference between reactants and products so as to maximize the reaction rate, we have $\epsilon = \lambda$ and the exponential dependence of the reaction rate on T_{eff} disappears; all we have left is $k \propto 1/\sqrt{T_{\text{eff}}}$, which indeed is a weak, inverse temperature dependence. This sort of fine tuning might make sense—perhaps evolution has selected for molecular parameters that maximize the electron transfer rates.

The structure if the reaction center is such that one can take out the quinone molecules and replace them with analogs that have different electron affinities, and in this way manipulate the value of ϵ. Perhaps surprisingly, increases in ϵ have very little effect on the rate constant for the recombination reaction $P^+Q^- \to PQ$, or on the forward reaction $I^-Q \to IQ^-$, and for all the values of ϵ probed one sees an approximately temperature independent rate. This argues strongly against tuning of $\epsilon = \lambda$ as an explanation for the observed "activationless" behavior.

---

there is no correction for the index of refraction. Once you start reading about molecular spectroscopy and chemical reactions (replete with calories and moles), you'll have to get some practice at changing units!

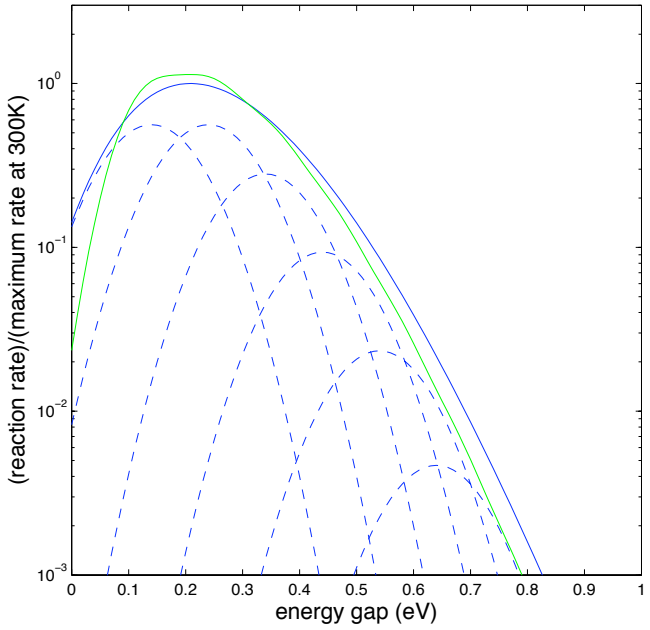

FIG. 42 Electron transfer coupled to high frequency vibrations, from Eq's (293) and (294). Dashed lines show contributions to the rate constant at $T = 300\,\mathrm{K}$ from processes that leave behind $n = 0, 1, 2, \ldots$ quanta in the high frequency mode. The total rate $k(T = 300\,\mathrm{K})$ is shown in the solid blue line, and $k(T = 30\,\mathrm{K})$ in green. The high frequency mode has $\hbar\Omega = 0.1\,\mathrm{eV}$ and $S = 1$.

Suppose that instead of one vibrational mode, we have two—one at a low frequency ω, which we can treat by the semi–classical argument given above, and one at a high frequency Ω that really needs a proper quantum mechanical description. The initial state of the high frequency mode is the ground state (since $k_BT \ll \hbar\Omega$), but in the final state we can excite one or more vibrational quanta, and the overall reaction rate will be a sum over terms corresponding to each of these possible final states. From the point of view of the low frequency mode, if the system transitions into a state with n high frequency quanta, this renormalizes the matrix element $\Delta \to \Delta_n$ and reduces the energy gap $\epsilon \to \epsilon - n\hbar\Omega$. Thus the rate constant becomes

$$k = \sum_{n=0}^{\infty} \frac{\Delta_n^2}{\hbar\sqrt{4\pi\lambda k_B T_{\mathrm{eff}}}} \exp\left[-\frac{(\epsilon - n\hbar\Omega - \lambda)^2}{4\lambda k_B T_{\mathrm{eff}}}\right], \quad (293)$$

where now λ refers only to the reorganization energy of the low frequency mode. Results are shown in Fig 42.

Problem 49: Renormalized matrix elements. To complete the calculation in Eq (293), we need to understand how the matrix elements are renormalized by coupling to the high frequency modes. Get the students to derive ...

$$\Delta_n^2 = e^{-S}\frac{S^n}{n!}\Delta^2, \quad (294)$$

and explain the meaning of S.

We see that the possibility of exciting different numbers of vibrational quanta greatly broadens the dependence of the rate constant on the energy gap ϵ, and provides a huge widening of the region over which we see very little (or even inverted) temperature dependence. This seems a more plausible and robust explanation of the observed activationless kinetics in the photosynthetic reaction center. Importantly, it relies in an essential way on the quantum behavior of the high frequency vibrational motions that are coupled to the electron transfer, and this is true even at room temperature. There is no shortage of such high frequency modes in the quinones, chlorophylls and pheophytins; what is interesting is the way in which the interplay of these quantum modes with the lower frequency classical modes (including, presumably, modes of the protein scaffolding itself) shapes the observed functional behavior.

A third issue is that, although we are talking about electron transfer reactions, we have said relatively little about the electrons themselves—there are two states, localized on the donor and acceptor sites, and there is a matrix element that connects these states, but that seems to be all. In fact we can say a bit more. First, our use of perturbation theory obviously depends on the matrix element not being too large. If we go back to our simple model of the DeVault–Chance reaction and try to fit the absolute rate constants as well as the temperature dependence, we find $\Delta \sim 10^{-4}\,\mathrm{eV}$. Certainly this is small compared with the other energies in the problem (λ, $\hbar\omega$, k_BT, ϵ), which indicates that our use of perturbation theory is consistent. [Finish the discussion of matrix elements!]

[Do we want to say anything about coherence and the very first, fastest steps??]

All other things being equal, quantum effects are stronger for lighter particles. As we have seen, electrons essentially always tunnel—there are almost no chemical or biochemical reactions involving thermal activation of an electron over a barrier. Since the early days of quantum mechanics, people have wondered if chemical reactions involving the next lightest particle, a proton or hydrogen atom, might also involve tunneling in a significant way. To be concrete, consider the situation in Fig 43, where the reaction coordinate is the position of the H atom itself, moving from donor to acceptor atom. But, while still attached to the donor atom (e.g., a carbon) we can observe vibrations of the D–H bond, and for C–H we know that the frequencies of these vibrations can be as high as $\nu \sim 2500 - 3000\,\mathrm{cm}^{-1}$. The vibrational quanta thus are $h\nu \sim 1/4 - 1/3\,\mathrm{eV}$. In fact the activation energies of many chemical reactions are not that much larger

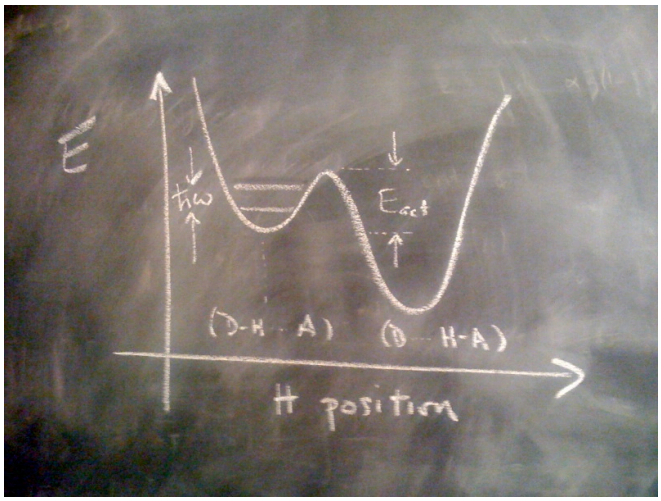

FIG. 43 Transfer of a hydrogen atom from a donor D to an acceptor A. The reaction coordinate is the position of the H atom, but we expect that quantization effects are non–negligible.

than this, perhaps $0.5 - 1\,\mathrm{eV}$. This means that, as indicated in the crude sketch, climbing up to the top of the barrier between reactants and products involves adding just two or three vibrational quanta. What this means is that the reaction can't really be completely classical, if the reaction coordinate really is the stretching of the bond itself.

If we make the crude approximation that the barrier is rectangular, with height E, then the rate of going over the barrier should be $k \propto e^{-E/k_BT}$, as before, while the rate of tunneling through the barrier is $k \propto e^{-2\sqrt{2mE}\ell/\hbar}$, where ℓ is the width of the barrier and m is the mass of tunneling particle. Although we could worry about the prefactors, the exponentials are probably the dominant effects, and so we might guess (as in the problem above) that tunneling is more important that classical thermal activation only if

$$e^{-E/k_BT} < e^{-2\sqrt{2mE}\ell/\hbar}, \tag{295}$$

or

$$T < T_0 \sim \frac{\hbar}{k_B}\sqrt{\frac{E}{2m\ell}}. \tag{296}$$

If the width of the barrier is $\ell \sim 1\,\text{Å}$, and its height is $E \sim 50\,\mathrm{kJ/mole}$, then with m the mass of the proton we find $T_0 \sim 190\,\mathrm{K}$, well below room temperature. Thus, although it might be difficult to see the transfer of a proton as being completely classical, it's also true that the transfer reaction is unlikely to be dominated by tunneling at room temperature if the barrier is static.

In the interior of a protein, we can imagine that the donor and acceptor are held by different parts of the large molecule, and as the protein flexes and breathes, these sites will move. Effectively this means that the width of the barrier will fluctuate. On average, this increases the probability of tunneling through the barrier. If the fluctuations in ℓ are Gaussian, the tunneling probability becomes

$$e^{-2\sqrt{2mE}\ell/\hbar} \to \left\langle e^{-2\sqrt{2mE}\ell/\hbar} \right\rangle = \exp\left[-2\sqrt{2mE}\bar{\ell}/\hbar + 4mE\langle(\delta\ell)^2\rangle/\hbar^2\right], \tag{297}$$

where $\bar{\ell}$ is the average width of the barrier and $\langle(\delta\ell)^2\rangle$ is the variance of the this width. With the parameters as before, the enhancement of the tunneling probability involves the term

$$4mE\langle(\delta\ell)^2\rangle/\hbar^2 \sim 6\left\langle \left(\frac{\delta\ell}{0.1\,\text{Å}}\right)^2 \right\rangle. \tag{298}$$

As described in Appendix A.5, measurements of Debye–Waller factors in X–ray diffraction from protein crystals provide estimates of the fluctuations in structure, and these structural fluctuations in are easily several tenths of an Ångström. Thus this term, which appears in the exponential, can be huge. This completely shifts the balance between tunneling and classical, thermal activation, so that in the presence of fluctuations it becomes plausible that tunneling is dominant at room temperature.

Notice that the role of protein vibrational motions here is very different than in the case of electron transfer. In electron transfer, there is a small matrix element that couples the two relevant states, and protein motions serve to bring these two states into degeneracy with one another. This effect presumably could happen in the case of proton transfer as well, but we have focused on the coupling of fluctuations to the tunneling matrix element. This coupling is especially interesting because it generates exponential terms in the reaction rate that have a dependence on mass ($\ln k \propto m$) that is very different from the naive tunneling exponent ($\ln k \propto -\sqrt{m}$) or the zero–point corrections to the activation energy ($\ln k \propto 1/\sqrt{m}$; see next problem); because this mass–dependent term also depends on the variance of structural fluctuations, it is also temperature dependent. Indeed, it was the discov-

ery of anomalous, temperature dependent isotope effects in enzymatic proton transfer reactions that prompted renewed discussion of these dynamical effects on tunneling.

Problem 50: Isotope effects. Chemical reaction rates change when we substitute one isotope for another. There is a "semiclassical" theory of these isotope effects, which says that the reaction proceeds by conventional thermal activation, but the activation energy is reduced by the zero–point energy of vibrations along the reaction coordinate, $k \propto \exp\left[-(E_{\rm act} - \hbar\omega)/k_B T\right]$.

(a.) Vibrational frequencies are proportional to $1/\sqrt{m}$, with m the (effective) mass of the particle(s) moving along the mode with frequency ω. In the simple picture where all of the motion along the reaction coordinate is dominated by the motion of the proton, derive a relationship between the ratios of rate constants for hydrogen, deuterium and tritium transfer.

(b.) If the reaction coordinate involves motion of atoms other than transferred hydrogen, what happens to the predicted magnitude of the isotope effects? What about the relationship you derived in [a.]?

(c.) [Let's do something with averaging over fluctuating barriers to see how isotope effects come out ...]

I hope that you take a few lessons away from this (long) discussion. First, chemical reactions are the result of fluctuations at the molecular level. We can describe the nature of these fluctuations in some detail, since rare events such as escape over a high barrier are dominated by specific trajectories. In large biological molecules, the flexibility of the molecule means that there is another way for fluctuations to be important, as the variations in protein structure, for example, couple to changes in the barrier for the relevant chemical rearrangements or bring weakly coupled electronic states into degeneracy. Finally, these fluctuations in protein structure can completely revise our view of whether the reaction itself proceeds via classical 'over the barrier' motion or by quantum tunneling. These theoretical observations, and the experiments to which they connect, suggest that Nature exploits not just the structure of biological molecules, but also the fluctuations in these structures, to control the rates of chemical reactions.

If you need a review of the Langevin equation, I like the treatment in the little book by Kittel (1958), as well as the somewhat longer discussion by Pathria (1972). Every physics student should understand the basic instanton calculation of tunneling, as an illustration of the power of path integrals. There is no better treatment than that given by Coleman in his justly famous Erice lectures. If you read Coleman you'll not only get a deeper view of what we have covered here, you'll get all the missing pieces about the prefactor of the rate constant, and much more. For more general background on path integrals, including some discussion of how to use them for classical stochastic processes, the standard reference is Feynman & Hibbs (1965). For more rigorous accounts of many of these issues (e.g., getting the Jacobian right in constructing the path integral), see Zinn–Justin (1989). The original discussion of diffusion (even with inertia) over a barrier is due to Kramers (1940); for a modern perspective see Hänggi et al (1990).

Coleman 1988: *Aspects of Symmetry* S Coleman (Cambridge University Press, Cambridge, 1988).

Feynman & Hibbs 1965: *Quantum Mechanics and Path Integrals* RP Feynman & AR Hibbs (McGraw–Hill, New York, 1965).

Hänggi et al 1990: Reaction–rate theory: Fifty years after Kramers. P Hänggi, P Talkner & M Borkovec, *Revs Mod Phys* **62,** 251–341 (1990).

Kittel 1958: *Elementary Statistical Physics* C Kittel (Wiley, New York, 1958).

Kramers 1940: Brownian motion in a field of force and the diffusion model of chemical reactions. HA Kramers *Physica* **7,** 284–304 (1940).

Pathria 1972: *Statistical Mechanics* RK Pathria (Pergamon Press, Oxford, 1972).

Zinn–Justin 1989: *Quantum Field Theory and Critical Phenomena* J Zinn–Justin (Oxford University Press, Oxford, 1989).

Myoglobin was the first protein whose structure was solved by X–ray diffraction. Aspects of X–ray analysis are described in Appendix A.5. For a perspective on myoglobin, see Kendrew (1964). The experiments on myoglobin are by Austin et al (1975), which touched off a huge followup literature. A clear discussion of the interplay between the a reaction coordinate and a protein coordinate was given by Agmon and Hopfield (1983). The demonstration of tunneling in this system is by Alberding et al (1976).

Agmon & Hopfield 1983: Transient kinetics of chemical reactions with bounded diffusion perpendicular to the reaction coordinate. N Agmon & JJ Hopfield, *J Chem Phys* **78,** 6947–6959 (1983).

Alberding et al 1976: Tunneling in ligand binding to heme proteins. N Alberding, RH Austin, KW Beeson, SS Chan, L Eisenstein, H Frauenfelder & TM Nordlund, *Science* **192,** 1002–1004 (1976).

Austin et al 1975: Dynamics of ligand binding to myoglobin. RH Austin, KW Beeson, L Eisenstein, H Frauenfelder & IC Gunsalus, *Biochemistry* **14,** 5355–5373 (1975).

Kendrew 1964: Myoglobin and the structure of proteins. JC Kendrew, in *Nobel Lectures in Chemistry 1942–1962* (Elsevier, Amsterdam, 1964). See also http://www.nobelprize.org.

Classical overviews of the photosynthetic reaction center are provided by Feher & Okamura (1978) and Okamura et al (1982). As with many biological molecules, many questions about the reaction center were sharpened once the structure was determined at atomic resolution (Deisenhoffer et al 1984); this work was important also as a demonstration that one could use the classical methods of X–ray crystallography (cf Appendix A.5) for proteins that are normally embedded in membranes. It should be emphasized, however, that the electron transfer reactions leave an enormous variety of spectroscopic signatures—separating charges not only changes optical properties of the molecules, it generates unpaired spins that can be seen using electron paramagnetic resonance (EPR), and the distribution of the spin across multiple atoms at the donor and acceptor sites can be mapped using electron–nuclear double resonance (ENDOR). An early view of the uses of EPR and ENDOR in biological systems is given by Feher (1970); this article appears in the proceedings of the first Les Houches physics summer school to be devoted to questions at the interface with biology [check this!]. For a synthesis of structural and spectroscopic data in relation to function, see Feher et al (1989).

DeVault & Chance 1966: Studies of photosynthesis using a pulsed laser. I. Temperature dependence of cytochrome oxidation in *Chromatium*. Evidence for tunneling. D DeVault & B Chance, *Biophys J* **6,** 825–847 (1966).

Deisenhoffer et al 1984: X–ray structure analysis of a membrane protein complex: Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodopseudomonas viridis*. J Deisenhoffer, O Epp, K Miki, R Huber & H Michel, *J Mol Biol* **180,** 385–398 (1984).

Feher 1970: Electron paramagnetic resonance with applications to selected problems in biology. G Feher, in *Physical Problems in Biological Systems*, C DeWitt & J Matricon, eds, pp 251–365 (Gordon & Breach, Paris, 1970).

Feher & Okamura 1978: Chemical composition and properties of reaction centers. G Feher & MY Okamura, in *The Photosynthetic Bacteria*, RK Clayton & WR Sistrom, eds, pp 349–386 (Plenum Press, New York, 1978).

Feher et al 1989: Primary processes in bacterial photosynthesis: structure and function of reaction centers. G Feher, JP Allen, MY Okamura & DC Rees, *Nature* **339,** 111–116 (1989).

Okamura et al 1982: Reaction centers. MY Okamura, G Feher & N Nelson, in *Photosynthesis: Energy Conversion by Plants and Bacteria, Volume 1*, Govindjee, ed, pp 195–272 (Academic Press, New York, 1982).

The original experiments that provided evidence for tunneling in photosynthetic electron transfer were done by DeVault and Chance (1966) on samples that were a bit messier than the purified reaction centers that emerged in subsequent years. The kinetics of the initial charge separation reactions were described by [fill in refs!]. The modern view of biological electron transfer reactions, including the role of tunneling in the vibrational degrees of freedom, is due to Hopfield (1974). Exploration of the energy gap dependence of reaction rates was pioneered by Gunner et al (1986), and the evidence for frozen distributions of electron transfer rates was provided by Kleinfeld et al (1984). For a review of efforts to calculate electronic matrix elements in real protein structures, see Onuchic et al (1992). [Maybe we need more here? Depends also on what happens in the text.]

DeVault & Chance 1966: Studies of photosynthesis using a pulsed laser. I. Temperature dependence of cytochrome oxidation in *Chromatium*. Evidence for tunneling. D DeVault & B Chance, *Biophys J* **6,** 825–847 (1966).

Gunner et al 1986: Kinetic studies on the reaction center protein from *Rhodopseudomonas sphaeroides*: the temperature and free energy dependence of electron transfer between various quinones in the Q_A site and the oxidized bacteriochlorophyll dimer. MR Gunner, DE Robertson & PL Dutton, *J Phys Chem* **90,** 3783–3795 (1986).

Hopfield 1974: Electron transfer between biological molecules by thermally activated tunneling. JJ Hopfield, *Proc Nat'l Acad Sci (USA)* **71,** 3640–3644 (1974).

Kleinfeld et al 1984: Electron–transfer kinetics in photosynthetic reaction centers cooled to cryogenic temperatures in charge–separated state: evidence for light–induced structural changes. D Kleinfeld, MY Okamura & G Feher, *Biochemistry* **23,** 5780–5786 (1984).

Onuchic et al 1992: Pathway analysis of protein electron–transfer reactions. JN Onuchic, DN Beratan, JR Winkler & HB Gray, *Annu Rev Biophys Biomol Struct* **21,** 349–377 (1992).

The papers that reignited interest in proton tunneling in enzymes were Cha et al (1989) and Grant & Klinman (1989). The idea that these experiments should be understood in terms of coupling between quantum motion of the proton and classical motion of the protein was developed by Bruno & Bialek (1992). It took roughly a decade for these ideas to solidify, as described in reviews by Sutcliffe & Scrutton (2002) and Knapp & Klinman (2002). [add refs to Nori et al in cytochrome oxidase?]

Bruno & Bialek 1992: Vibrationally enhanced tunneling as a mechanism for enzymatic hydrogen transfer. WJ Bruno & W Bialek, *Biophys J* **63,** 689–699 (1992).

Cha et al 1989: Hydrogen tunneling in enzyme reactions. Y Cha, CJ Murray & JP Klinman, *Science* **243,** 1325–1330 (1989).

Grant & Klinman 1989: Evidence that both protium and deuterium undergo significant tunneling in the reaction catalyzed by bovine serum amine oxidase. KL Grant & JP Klinman, *Biochemistry* **28,** 6597–6695 (1989).

Knapp & Klinman 2002: Environmentally coupled hydrogen tunneling: Linking catalysis to dynamics. MJ Knapp & JP Klinman, *Eur J Biochem* **269,** 3113–3121 (2002).

Sutcliffe & Scrutton 2002: A new conceptual framework for enzyme catalysis: Hydrogen tunneling coupled to enzyme dynamics in flavoprotein and qunioprotein enzymes. MJ Sutcliffe & NS Scrutton, *Eur J Biochem* **269,** 3096–3102 (2002).

B. Molecule counting

Many of the crucial signals in biological systems—signals that are internal to cells, signals that cells use to communicate with one another, even signals that organisms exchange—are carried by changes in the concentration of specific molecules. The molecules range in size from single ions (e.g., calcium) to whole proteins. Such chemical signals act by binding to specific targets, whose synthesis and accessibility can also be controlled by the cell. A key point is that individual molecules move randomly, and so the arrival of signals at their targets has some minimum level of noise. As we shall see, several different systems operate with a reliability close to this physical limit: in essence, these systems are counting every molecule, and making every molecule count.

In what follows we will see examples of chemical signaling in the decisions that cells make about whether to read out the information encoded in particular genes, in the trajectories that axons take toward their targets in the developing brain, in the control signals that bacteria use to regulate their movement, and in the development of spatial patterns in a developing embryo. But much of our thinking about precision, reliability and noise in chemical signaling has been shaped by the phenomena of chemotaxis in bacteria, so this is where we will start.

Although our experience with other animals makes it clear that we are not alone in our ability to sense the world, it still seems remarkable that single celled organisms such as bacteria are endowed with sensory systems

that allow them to move in response to a variety of signals from the environment, including the concentrations of various chemicals. A classical observation (from the 19th century) is that some bacteria, swimming in water on a microscope slide, under a cover slip, will collect at the center of cover slip, while others will collect at the edges. Those with more refined tastes will form a tight band that traces the outlines of the square cover slip. Oxygen diffuses into the water through the edges of the cover skip, and by collecting along a square the bacteria have migrated to a place of constant (not maximal or minimal) oxygen concentration. It is plausible that this happens because they can sense the oxygen concentration and "know" the most comfortable value of this concentration, much as we might move to be the most comfortable distance from a fireplace in an otherwise unheated room.

That bacteria collect at nontrivial concentrations of different molecules really doesn't demonstrate that they sense the concentration. They might instead sense some internal consequences of the external variables, such as the accumulation of metabolic intermediates. In the 1960s Adler found mutants of *E coli* which cannot metabolize certain sugars or amino acids but will nevertheless migrate toward the sources of these molecules; also there are mutants that metabolize but can't migrate. This is convincing evidence that metabolism and sensing are separate systems, and thus begins the fruitful exploration of the sensory mechanisms of bacteria and the connection of these sensory mechanisms to motor output. This phenomenon is called *chemotaxis.*

I'll skip lots of the truly classical stuff and proceed with the modern biophysical approach, which begins $\sim$ 1970. To a large extent this modern approach rests on the work of Howard Berg and collaborators. The first key step taken by Berg and Brown was to observe the behavior of individual bacteria. *E coli* are $\sim 1\,\mu$m in size, and can be seen relatively easily under the light microscope, but since the bacteria swim at ~ 20 body lengths per second they easily leave the field of view or the plane of focus; the solution is to build a tracking microscope.

Observations in the tracking microscope, as in Fig 44, showed that the trajectories of individual bacteria consist of relatively straight segments interrupted by short intervals of erratic "motion in place." These have come to be called runs and tumbles, respectively. Tumbles last ~ 0.1 seconds, but the erratic motion during this brief time is sufficient to cause successive runs to be in almost random relative directions. Thus the bacterium runs in one direction, then tumbles and chooses a new direction at random, and so on. Runs themselves are distributed in length, as if the termination of a run is itself a random process.

Closer examination of the runs shows how it is possible for this seemingly random motion to generate progress up the gradient of attractive chemicals. When the bacterium runs up the gradient, the mean duration of the runs becomes longer, biasing the otherwise random walk. Interestingly, when bacteria swim down the gradient (of an attractant, or up the gradient of a repellent) the is relatively little change in the mean run length. Berg has described this as a form of optimism: If things are getting better, keep going, but if things are getting worse, don't worry. [Need to look at the notion of optimism once more in relation to all the data.]

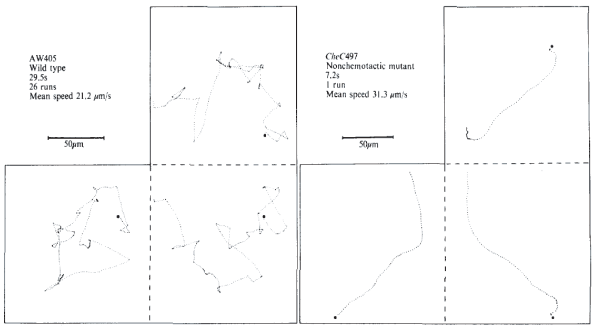

FIG. 44 Paths of *E coli* as seen in the original tracking microscope experiments, from Berg & Brown (1972). The three panels in each case are projections of the path onto the three orthogonal planes (imagine folding the paper into a cube along the dashed lines). At left, wild type bacteria, showing the characteristic runs and tumbles. At right, a non–chemotactic mutant that never manages to tumble.

Since runs get longer when bacteria swim along a positive gradient, it is natural to ask whether the cell is responding to the spatial gradient itself or to the change in concentration with time along the path. As we will see, the spatial gradients to which the cell can respond are very small, and searching for a systematic difference (for example) between the front and back of the bacterium is unlikely to be effective just on physical grounds, independent of biological mechanisms. Indeed, this is the reason why chemotaxis is such an important example of the issues in this section. To search for a time domain mechanism one can expose the bacteria to concentrations which are spatially uniform but varying in time; if the sign of the change corresponds to swimming up a positive gradient, runs should be prolonged. The first such experiment used very large, sudden changes in concentration, and found that cells were trapped in extremely long runs. A more sophisticated experiment used enzymes to synthesize attractants from inert precursors, exposing the cells to gradual changes more typical of those encountered while swimming. Purely time domain stimuli were sufficient to generate modulations of run length that agree quantitatively with those observed for bacteria experiencing spatial gradients.

Problem 51: Chemotaxis in one dimension. To make the intuition of the previous paragraphs more rigorous, consider a simplified problem of chemotaxis in one dimension. There are then two populations of bacteria, the + population that moves to the right and the − population that moves to the left, each at speed v. Let the probability of finding a + [−] bacterium at position x be $P_+(x,t)$ [$P_-(x,t)$]. Assume that the rate of tumbling depends on the time derivative of the concentration along the bacterial trajectory as some function $r(\dot{c})$, where for the ± bacteria we have $\dot{c} = \pm v dc/dx$, and that cells emerge from a tumble going randomly left or right.

(a.) Show that the dynamics of the two probabilities obey

$$\frac{\partial P_+(x,t)}{\partial t} + v\frac{\partial P_+(x,t)}{\partial x} = -r\left(v\frac{dc}{dx}\right)P_+(x,t) + \frac{1}{2}\left[r\left(v\frac{dc}{dx}\right)P_+(x,t) + r\left(-v\frac{dc}{dx}\right)P_-(x,t)\right] \tag{299}$$

$$\frac{\partial P_-(x,t)}{\partial t} - v\frac{\partial P_-(x,t)}{\partial x} = -r\left(-v\frac{dc}{dx}\right)P_+(x,t) + \frac{1}{2}\left[r\left(v\frac{dc}{dx}\right)P_+(x,t) + r\left(-v\frac{dc}{dx}\right)P_-(x,t)\right]. \tag{300}$$

Explain the meaning of each of the terms in terms of what happens as cells enter into and emerge from tumbles. Note that in this approximation tumbles themselves are instantaneous, which isn't so bad (0.1 s vs the $\sim 1-10$ s for typical runs).

(b.) To see if the bacteria really migrate toward high concentrations, look for the steady state of these equations. If we simplify and assume that the rate of tumbling is modulated linearly by the time derivative of the concentration,

$$r(\dot{c}) \approx r(0) + \frac{\partial r}{\partial \dot{c}}\dot{c} + \cdots, \tag{301}$$

show that

$$P(x) = \frac{1}{Z}\exp\left[-\frac{\partial r}{\partial \dot{c}}c(x)\right]. \tag{302}$$

Thus, in these approximations, chemotaxis leads to a Boltzmann distribution of bacteria, in which the concentration acts as a potential. If the molecules are attractive then $\partial r/\partial \dot{c} < 0$ and hence maxima of concentration are minima of the potential, conversely for repellents. The stronger the modulation of the tumbling rate (as long as we stay in our linear approximation) the lower the effective temperature and the tighter the concentration of bacteria around the local maxima of concentration.

Problem 52: Nonlinearities. Within this simplified one dimensional world, can you make progress without the approximation that $r(\dot{c})$ is linear? More specifically, what is the form of the stationary distribution $P(x)$ that solves Eq (**??**) for nonlinear $r(\dot{c})$? Can you show that there still is an effective potential with minima located at places where the concentration is maximal?

Problem 53: A little more about the effectiveness of chemotaxis.

(a.) Within the one dimensional model, what happens if the tumbling rate is modulated not just by the time derivative, but also by the absolute concentration, so that the bacterium confuses "currently good" for "getting better"?

(b.) Can you generalize this discussion to three dimensions? Instead of having just two groups + and −, one now needs a continuous distribution $P(\mathbf{\Omega}, x, t)$, where $\mathbf{\Omega}$ denotes the direction of swimming. Derive an equation for the dynamics of $P(\mathbf{\Omega}, x, t)$ in the same approximations used above, and see if the Boltzmann–like solution obtains in this more realistic case.

All of this description so far is about the phenomenology of swimming. But how does it actually work? The basic problem is that bacteria are too small to take advantage of inertia. When we swim, we can push off the wall of the pool and glide for some distance, even without moving our arms or legs; this gliding distance is on the order of one or two meters, roughly the length of our bodies. In contrast, if a bacterium stops running its motors, it will glide for a distance comparable not to its body length ($\sim 1\,\mu$m) but to the diameter of an atom. To see this, think about a small particle moving through a fluid, subject only to drag forces (the motors are off). If the velocities are small, we know the drag will be proportional to the velocity, so Newton's equation is just

$$m\frac{dv}{dt} = -\gamma v. \tag{303}$$

For a spherical object or radius r, Stokes' law tells us that $\gamma = 6\pi\eta r$, where η is the viscosity of the fluid, and we also know that $m = 4\pi\rho r^3/3$, where ρ is the density of the object. The result is that

$$v(t) = v(0)\exp(-t/\tau), \tag{304}$$

where

$$\tau = \frac{m}{\gamma} = \frac{2\rho r^2}{9\eta}. \tag{305}$$

If we assume that the density of bacteria is roughly that of water, then it is useful to recall that η/ρ has units of a diffusion constant, and for water $\eta/\rho = 0.01\,\mathrm{cm}^2/\mathrm{s}$. With $r \sim 1\,\mu\mathrm{m} = 10^{-4}\,\mathrm{cm}$, this gives $\tau \sim 5\times 10^{-7}\,\mathrm{s}$. If the initial velocity is $v(0) \sim 20\,\mu$m/s, the net displacement during this coasting is $\Delta x = v(0)\tau \sim 10^{-11}\,\mathrm{m}$; recall that a hydrogen atom has a diameter of $\sim 1\,\text{Å} = 10^{-10}\,\mathrm{m}$.

The conclusion from such simple estimates is that bacteria can't coast. More generally, mechanics on the scale of bacteria is such that inertia is negligible, as if Aristotle (rather than Galileo and Newton) were right. This really about the nature of fluid flow on this scale.[36] For an incompressible fluid (which is a good approximation here—surely the bacteria don't generate sound waves as

[36] My experience is that most physics students don't know too much fluid mechanics, so although this is elementary I put it here. For a more thorough discussion, see, as usual, Landau and Lifshitz.

they swim), the Navier–Stokes equations are

$$\rho\left[\frac{\partial \mathbf{v}}{\partial t}+\mathbf{v}\cdot\nabla\mathbf{v}\right]=-\nabla p+\eta\nabla^2\mathbf{v}, \tag{306}$$

where $\mathbf{v}$ is the local velocity of the fluid, p is the pressure, and as usual ρ is the density and η is the viscosity. The pressure isn't really an independent variable, but needs to be there so we can enforce the condition of incompressibility,

$$\nabla\cdot\mathbf{v}=0. \tag{307}$$

These equations need to be supplemented by boundary conditions, in particular that the fluid moves with the same velocity as any object at the points where it touches that object. Thus the velocity should be zero at a stationary wall, and should be equal to the velocity of a swimmer at the swimmer's surface.

Problem 54: Understanding Navier–Stokes. This isn't a fluid mechanics course, but you should be sure you understand what Eq (306) is saying. In particular, this is nothing but Newton's $F=ma$. Explain.

Dimensional analysis is an enormously powerful tool in fluid mechanics. We are free to choose new units for length (ℓ) and time (t_0), and hence for velocity ($v_0=\ell/t_0$), as well as for pressure p_0, and this gives us

$$\rho\left[\frac{v_0}{t_0}\frac{\partial\tilde{\mathbf{v}}}{\partial\tilde{t}}+\frac{v_0^2}{\ell}\tilde{\mathbf{v}}\cdot\tilde{\nabla}\tilde{\mathbf{v}}\right]=-\frac{p_0}{\ell}\tilde{\nabla}\tilde{p}+\eta\frac{v_0}{\ell^2}\tilde{\nabla}^2\tilde{\mathbf{v}}, \tag{308}$$

$$\frac{\rho\ell v_0}{\eta}\left[\frac{\partial\tilde{\mathbf{v}}}{\partial\tilde{t}}+\tilde{\mathbf{v}}\cdot\tilde{\nabla}\tilde{\mathbf{v}}\right]=-\frac{p_0\ell}{\eta v_0}\tilde{\nabla}\tilde{p}+\tilde{\nabla}^2\tilde{\mathbf{v}}, \tag{309}$$

where $\tilde{t}=t/t_0$, $\tilde{\mathbf{v}}=\mathbf{v}/v_0$, and $\tilde{p}=p/p_0$. Now we can set $p_0\ell/\eta v_0=1$, which gets rid of all the units, except we are left with a dimensionless combination

$$\mathrm{Re}\equiv\frac{\rho\ell v_0}{\eta} \tag{310}$$

which is called the Reynolds' number. Notice that if we choose the unit of length to be the size of the objects that we are interested in, and v_0 to be the speed at which they are moving, then even the boundary conditions don't have any units, nor do they introduce any dimensionless factors that are far from unity. The conclusion is that all fluid mechanics problems with the same geometry (shapes) are the same if they have they have the same Reynolds' number. In this sense, being smaller (reducing ℓ) is the same as living at increased viscosity.[37]

To make a long story short, we live at high Reynolds' number, and bacteria live at low Reynolds' number (Fig 45). Turbulence is a high Reynolds' number phenomenon, as is the more mundane gliding through the pool after we push off the wall. At low Reynolds' number, life is very different. Inertia is absent, and so forces must balance at every instant of time. To say this more startlingly, if $\mathrm{Re}\to 0$ then time doesn't actually appear in the equations. This means that, as you swim, the distance that you move depends on the *sequence* of motions that you go through, but not on the dynamics with which you execute them.

To use Purcell's evocative example, at high Reynolds' number a scallop can propel itself by snapping shut, expelling a jet of water, and then opening slowly.[38] The jet will propel the scallop forward, and the drag of reopening can be made small by moving slowly. At low Reynolds' number this doesn't work, and the forward displacement generated by snapping shut will be exactly compensated by the drag on reopening. To have net movement from a cycle, the *sequence* of shapes that the swimmer goes through in the cycle must break time reversal invariance,

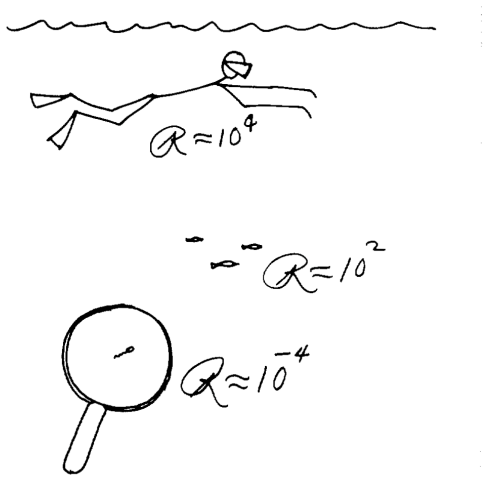

FIG. 45 Purcell's delightful sketch, illustrating the range of Reynolds' numbers relevant for swimming in humans, fish, and bacteria. From Purcell (1977).

---

[37] It is worth reflecting on the level of universality that we have here. We could imagine starting with a molecular description of fluids, then figuring out that, on the relevant length and time scales, all we need to know are the density and viscosity. Now we see that even these quantities are tied up with our choice of units. If we want to know what happens in natural units (i.e., scaling to the size and speed of the objects we are looking at), then all that matters is a single dimensionless combination, Re.

[38] There is an interesting issue about what real scallops do. Check Rob's note about this!

not just the trajectory.

So, how do bacteria evade the "scallop theorem"? If you watch them swimming, you can see that the have long filaments sticking out, and these seem to be waving. I emphasize that "see" is tough here. [This needs pictures; check with Berg.] These filaments are very small, $\sim 20\,\mathrm{nm}$ in diameter, much thinner than the wavelength of light. To see them, the easiest thing is to use dark field microscopy, in which the sample is illuminated from the side and what you see is the light scattered by $\sim 90^\circ$. These apparently waving appendages are called flagella, and remind us of what we see on other small swimming cells, such as sperm. The difference is that the flagella in these other cases are huge by comparison with the bacterial flagella. If you slice through the tail of a sperm and take an electron micrograph, you find an enormously complex structure, and if you try to analyze the system biochemically you find it is made from many different proteins. Importantly, some of these proteins act as enzymes and eat ATP, which we know is a source of energy, for example in our muscles. In contrast, the bacterial flagellum is small, with a relatively simple structure, and the biochemistry suggests that it is little more than a very long polymer made from one kind of protein; this protein is not an enzyme. How can this simple structure, with no ATPase activity, generate motions?

In experiments that aimed at better ways to see the flagella, one can attach "flags" to them using viruses that would stick to the flagella via antibodies. Once in a while, a virus with antibodies on both ends would stick to two flagella from different bacteria. When this happened, you could see the bacterial cells rotating, which one can imagine was a huge surprise. Eventually people figured out how to break off the flagella and stick the bacteria to a glass slide by the remaining stump, and then the bacterium rotates. Rotation can look like a wave if the flagellum is shaped like a corkscrew, and it is. Rotating a corkscrew obviously violates time reversal invariance. If you have several corkscrews and you rotate them with the correct handedness, they can fit together into a bundle. If you rotate the other way, the corkscrews clash, and any bundle will be blown apart by this clashing. So, with many flagella projecting from their surface, we can imagine that by switching the direction of rotation, the bacterium switches between a bundle that can smoothly propel the cell forward, and many independently moving flagella that would cause the cell to tumble in place—runs and tumbles correspond to counterclockwise and clockwise flagellar rotation.[39] If you find mutants that never tumble, and stick them down by their stumps, then they all rotate one way; similarly, mutants that tumble too often rotate the other way.

There is much more to say about the rotary engine itself, sitting at the base of the flagella. It is powered not by ATP but by a difference in chemical potential for hydrogen ions between the inside and the outside of the cell. This is an energy source that all cells use, albeit in different ways, because it allows chemical events at very different spatial locations to be coupled. Thus, as described in the preceding section, photosynthetic organisms use the energy of the absorbed photons to move electrons across a membrane, and then compensate the charges by moving protons; the resulting difference in chemical potential can be used by other membrane–spanning enzymes to make ATP, without being anywhere near the molecules that absorb the photon.[40] In fact, these enzymes that synthesize ATP also rotate as they let protons move down the gradient in their chemical potential, and these same enzymes are responsible for ATP synthesis in all cells. So, proton driven rotary motors are at the heart of energy conversion in all organisms.

There is also more to say about mechanics at low Reynolds' number. Swimming involves changing shape, and this provides the boundary conditions on the Navier–Stokes equations. A cycle of changing boundary conditions should lead to a net displacement. There is some subtlety here, since the space of shapes is not so easy to parameterize. If we think, for example, about a closed surface, "shape" is defined by three dimensional position as a function of the two coordinates on the surface (e.g., latitude and longitude), but there is an arbitrariness in how we choose these coordinates; of course any physical quantity, such as the amount by which the swimmer moves forward, must be invariant to this choice. Looking more closely, the freedom to choose coordinates means that the natural formulation of the problem includes a gauge symmetry. Reluctantly, let's leave all this and go back to the problem of chemotaxis itself.

Problem 55: Switching in tethered bacteria. As noted above, one way of studying bacterial motility and chemotaxis is to "tether" a bacterium by the stump of one flagellum, observing the rotation of the whole cell rather than the rotation of the flagellum. The file `omega.txt` contains a very long time series of the angular velocity from such an experiment done by WS Ryu, now at the University of Toronto.[41] The samples are taken sixty times per second, and the units of velocity are not quite arbitrary but

---

[39] This association goes of course depends on our convention for defining the handedness of rotation; it doesn't matter (and I have trouble remembering it!) as long as you are consistent.

[40] You can imagine how confusing this was before people figured it out! It looked like a mysterious action at a distance.

[41] Data that you need can be found at http://www.princeton.edu/∼wbialek/PHY562/data.html. [What is the permanent way of dealing with this??]

not really important either; you should be able to load this into MATLAB (`load omega.txt`).

(a.) You should see that the velocity switches between positive and negative values, but these values are fairly constant. This is consistent with swimming by switching between runs and tumbles, with little or no modulation of the swimming speed. What is the distribution of times spent with during each segment of positive or negative (clockwise or counterclockwise) velocity?

(b.) It usually is said that switching is a Poisson process, so that (as you remember from the discussion of photon counting) the distribution of intervals between switches should be exponential. Are your results in [a] consistent with this prediction?

(c.) Look carefully at the velocity vs. time in the data set. Are the data statistically stationary (time–translation invariant)? If you focus on segments of the data that are more clearly stationary, does that change your conclusions in [b]?

(d.) Sometimes the angular velocity makes a "partial switch," a brief excursion away from the typical positive or negative value but not quite a full switch to the opposite direction of rotation. Qualitatively, what is happening in these cases? What would be the simplest model to describe the velocity vs. time during such an event? Can you give a quantitative analysis of the data, fitting to your model? This is a bit open ended.

We are interested in the question of how sensitively the bacterium can respond to small concentration gradients. We suspect that, since individual molecular motions are random, there must be a limit, analogous to the shot noise in counting photons. In a classic paper, Berg and Purcell provided a clear intuitive picture of the noise in 'measuring' chemical concentrations. Their argument, schematized in Fig 46, was that if we have a sensor with linear dimensions a, then effectively the sensor samples a volume a^3. In this volume we expect to count an average of $N \sim ca^3$ molecules when the concentration is c. Each

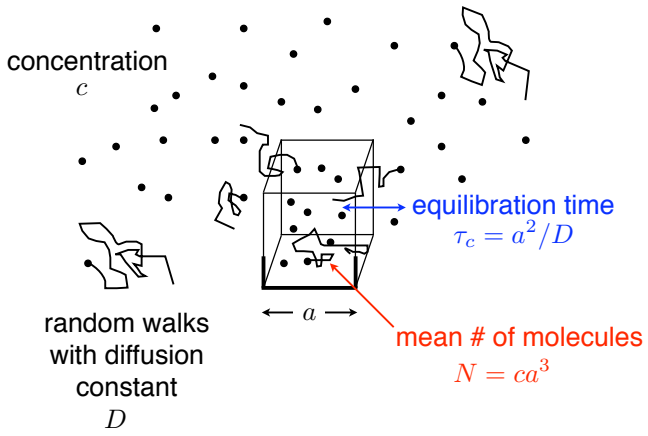

FIG. 46 A schematic of concentration measurements. A receptor of linear dimension a samples a volume a^3 and hence sees a mean number of molecules $N = ca^3$, where c is the concentration. These molecules random walk in and out of the sensitive volume with a diffusion constant D, corresponding to an equilibration or correlation time $\tau_c = a^2/D$.

such measurement, however, is associated with a noise $\delta N_1 \sim \sqrt{N}$. Since the count of molecules is proportional to our estimate of the concentration, the fractional error will be the same, so from one observation we obtain a precision

$$\left.\frac{\delta c}{c}\right|_1 = \frac{\delta N_1}{N} = \frac{1}{\sqrt{N}} = \frac{1}{\sqrt{ca^3}}. \tag{311}$$

We can make more accurate measurements by averaging over time, although this is a bit tricky—we won't get a better estimate of the concentration around us by counting the same molecules over and over again. Thus if we are willing to average over a time $\tau_{\rm avg}$, we can make K independent measurements, where $K \sim \tau_{\rm avg}/\tau_c$, and the correlation time τ_c is the time we have to wait in order to get an independent sample of molecules.

How do we get independent samples? If we look in a small volume, the molecules that we are looking at exchange with the surroundings through diffusion. Thus the time required to get an independent collection of molecules is the time required for molecules to diffuse in and out of the volume, $\tau_c \sim a^2/D$. Putting everything together we have

$$\frac{\delta c}{c} = \frac{1}{\sqrt{K}} \cdot \left.\frac{\delta c}{c}\right|_1 \tag{312}$$

$$= \sqrt{\frac{\tau_c}{\tau_{\rm avg}}} \cdot \frac{1}{\sqrt{ca^3}} \tag{313}$$

$$= \sqrt{\frac{a^2}{D\tau_{\rm avg}}} \cdot \frac{1}{\sqrt{ca^3}} \tag{314}$$

$$= \frac{1}{\sqrt{Dac\tau_{\rm avg}}}. \tag{315}$$

This is a lovely result. It says that the limit to the accuracy of measurements depends on the absolute concentration (more molecules → more accuracy), on the size the detector (bigger detectors → more accuracy), on the time over which we are willing to average (more time → more accuracy), and finally on the diffusion constant of the molecules we are sensing, because faster diffusion lets us see more independent samples in the same amount of time. All these parameters combine simply, essentially in the only way allowed by dimensional analysis.

One way of understanding this result on the limits to precision is to think about the rate at which molecules find their target. For molecules at concentration c moving with diffusion constant D, the rate (number of molecules per second) that arrive at a target of size a should be proportional both to c and to D, and then by dimensional analysis we need one factor of length, so the rate is $\sim Dac$ molecules per second. This result is used most often to talk about the "diffusion limited rate constant" for a chemical reaction; if we have

$$A + B \stackrel{k_+}{\to} C, \tag{316}$$

then the second order rate constant k_+ can never be bigger than $\sim Da$, where D is the diffusion constant of the molecules and a is their size, or more precisely the size of the region where they have to hit in order to react. But if the rate of molecular arrivals is $\sim Dac$, in a time $\tau_{\rm avg}$ we will count $\sim Dac\tau_{\rm avg}$ molecules, and if these molecules are arriving at random then there will be the usual square root fluctuations, which leads us to Eq (315). In this view, the Berg–Purcell limit is nothing but shot noise in molecular arrivals, and thus is completely analogous to shot noise in photon arrivals. Photons propagate and molecules diffuse, but under most conditions they both arrive at random, hence there is shot noise in counting.

Problem 56: Diffusion limited rates, more carefully. One can try a more careful calculation of the rate at which molecules find their target by diffusion. Image a sphere of radius a such that all molecules which hit the surface are immediately absorbed. Outside the sphere, the concentration profile must obey the diffusion equation, and the absorbtion means that on the spherical surface the concentration will be zero. Far from the sphere, the concentration should be equal to c. Thus we have

$$\frac{\partial c(\mathbf{x},t)}{\partial t} = D\nabla^2 c(\mathbf{x},t); \tag{317}$$

$$c(|\mathbf{x}| = a, t) = 0, \tag{318}$$

$$c(\mathbf{x}\to\infty, t) = c. \tag{319}$$

The number of molecules arriving per second at the surface of the sphere is given by an integral of the diffusive flux over the surface

$$\text{rate} = \int d^2s\, \hat{\mathbf{n}}\cdot\left[-D\nabla c(\mathbf{x},t)\right]\Big|_{|\mathbf{x}|=a}, \tag{320}$$

where d^2s is an element of the surface area on the sphere, and $\hat{\mathbf{n}}$ is the unit vector normal to the sphere.

(a.) Solve Eq (317), with the boundary conditions in Eqs (318) & (319), in steady state. Note that as a first step you should go to spherical coordinates; recall that in three dimensions the Laplacian can be written as

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\left[\frac{1}{\sin^2\phi}\frac{\partial^2}{\partial\theta^2} + \frac{1}{\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial}{\partial\phi}\right)\right], \tag{321}$$

where as usual r is the radius and θ and ϕ are the polar and azimuthal angles, respectively.

(b.) Use your steady state solution to evaluate the rate at which molecules arrive at the sphere, using Eq (320). Also, explain why simple dimensional analysis of these equations yields rate $\sim Dac$.

(c.) What happens if you try to give a dimensional analysis argument for the rate in one or two dimensions? If there are problems, can you explain how these problems either go away or are made more precise by trying to solve the diffusion equation with appropriate boundary conditions? As a hint, the two dimensional case is a bit delicate; focus first on one dimension.

Bacteria such as *E coli* have been observed to perform chemotaxis in environments where ambient concentrations of attractants such as sugars or amino acids are as low as $\sim 1\,$nM, which is $\sim 10^{-9}\times(6\times10^{23})/10^3 = 6\times10^{11}$ molecules per cm^3. These small molecules diffuse through aqueous solution with $D\sim 10^{-5}\,$cm^2/s, and the most generous assumption would be that the relevant size of the detector is the size of the whole bacterium, $a\sim 1\,\mu$m. Putting these factor together, we have $Dac\sim 600\,$s^{-1}. Thus, if the bacterium integrates for $\tau_{\rm avg}\sim 1.5\,$s, the smallest concentration changes it can detect are $\delta c/c\sim 1/30$. If the cells were to detect the difference in concentrations across the $\sim 1\,\mu$m length of their body, this would mean that the concentration was varying significantly on the scale of $30\,\mu$m, which is very short indeed. In real experiments (and, presumably, in the natural environment) the length scales of concentration gradients are one to two orders of magnitude longer. Thus, it's impossible—without integrating for minutes or hours—for bacteria to perform as they do by measuring a spatial gradient. The only possibility is to measure the concentration variation in time, along the trajectory that the bacterium takes through the gradient. Since the cells move at $v = 10-20\,\mu$m/s, on times scales of $\tau_{\rm avg}\sim 1.5\,$s this increases the signal by a factor of ten to thirty, and brings the signal above the background of noise, allowing for reliable detection.

[Maybe add remarks that this argument still works at higher concentrations, if the length scales of gradients are even longer? Perhaps this could be put into a problem?]

Although the comparisons are a bit rough,[42] we can draw several conclusions. First, real bacteria perform chemotaxis in response to small signals with a reliability close to the limits set by the physics of diffusion. Second, this is possible only if the cell measures the derivative of concentration vs. time as it moves, not spatial gradients across its body. Finally, to reach a reasonable signal–to–noise ratio requires that the cell average over time for more than one second.

Why don't the bacteria integrate for longer, and reduce the noise further? If you look closely at the trajectories of the bacteria, you can see that the longer runs curve a bit. In fact, the bacteria are sufficiently small that their own rotational Brownian motion disorients them on a time scale of ten or fifteen seconds. So, if you integrate for longer than this, you are no longer integrating something related to the gradient in a particular direction, or even your current direction of motion. This suggests that there is a physical limit setting the longest useful integration time.

Berg and Purcell also argued that there is a minimum

---

[42] I think there is an opportunity for a better experiment here. One could imagine analyzing the moments of transition from run to tumble (and back) in the same way that we analyze the action potentials from sensory neurons (see Section II.C), measuring the reliability of discrimination between small differences in concentration or reconstructing the concentration vs. time along the trajectory of a freely swimming bacterium.

useful integration time. Recall that molecules moving via diffusion traverse a distance $x_{\rm diff} \sim \sqrt{Dt}$ in a time t; in contrast, swimming at velocity v moves the bacterium by a distance $x_{\rm swim} \sim vt$. For short times, diffusion, with its square root dependence on time, goes farther than ballistic swimming motion. This means that on short time scales, the molecules that the bacterium sees along its path are the same molecules, and hence it really isn't combining statistically independent measurements. So, there is a minimum useful integration time (assuming you want to improve the signal–to–noise ratio by integrating) of $\tau \sim D/v^2$, and this works out to be about one second. Put in a pointer to a problem in the next section.

So, the strategy of *E coli* for measuring gradients is incredibly constrained by physics. To reach the observed performance, it has to count nearly every molecule that arrives at its surface. Even with this near ideal behavior, it can work only by making comparsions across time, not space, and estimates of time derivatives have to be averaged for a few seconds, not more and not less. This set of predictions about chemotactic strategy is almost parameter free, even if not precisely quantitative.

What do real bacteria do? We have already seen that they make temporal comparisons. Does the detailed form of these comparisons agree with the Berg–Purcell predictions? Although one could probably do better with modern experimental techniques, the best test was done in the early 1980s. In these experiments, bacteria were tethered to a glass slide and exposed to changing concentrations of attractants or repellents; a long series of such observations is then combined to measure the probability that the flagellar motor is rotating counterclockwise (corresponding to running) as function of time relative to the changing concentration. A summary of these experiments is shown in Fig 47. We see that the probability of running is modulated by the time derivative of the concentration, averaged over a window of a few seconds, exactly as predicted by the Berg–Purcell argument.

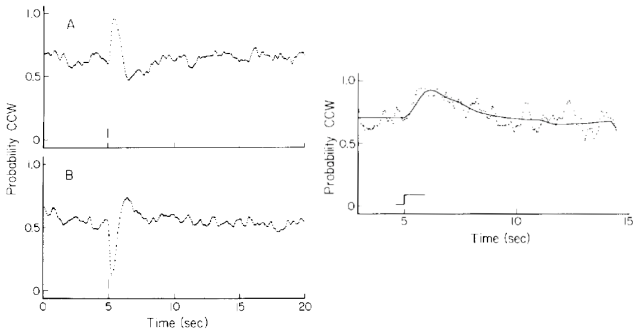

FIG. 47 Impulse responses in bacterial chemotaxis, from Block et al (1982). At left, changes in the probability of counterclockwise rotation of the motor, corresponding to running, as a function of time in response to a pulse of attractant (top) or repellent (bottom). We see that the form of the response is equivalent to integrating the time derivative of the input over a window or severa seconds. At right, the response to a step of attractant again has the form expected if we integrate the derivative over a short window. The real data are compared with a prediction based on integrating the response to impulses shown at left, and the agreement is good, as if the system were linear.

Being sensitive to a derivative means that the response to a step comes back almost exactly to the baseline before the step, as seen at right in Fig 47, so that the constant signal is ignored at times long after it was turned on. This gradual 'forgetting' of a constant signal is common in biological systems, and such phenomena are called 'adaptation.' All of our sensory systems exhibit adaptation, the most familar being the experience of stepping into a dark movie theater or out into the bright sunlight; at first we are acutely aware of the large difference in overall light intensity, but after a while everything looks normal and we are insensitive to the absolute photon flux. The case of bacteria is interesting because it seems that the adaptation is nearly perfect.

Experiments of the sort pictured in Fig 47 also make it possible to estimate the absolute sensitivity of the system in perhaps more compelling units. [should put the numbers here, maybe reproduce a figure] We now know how many receptors there are on the cell's surface, and so we can convert changes in concentration into changes in the number of occupied receptors. Indeed, one extra occupied receptor leads to a significant change in the probability of running vs tumbling. So, as expected, the bacterium is responding to individual molecular events.

This all seems a great success: much of bacterial behavior is understandable, semi–quantitatively, as a response to the physical constraints posed by life at low Reynolds' number and the noise in molecular counting; one can go further and say that bacterial behavior is near optimal in relation to this noise. On the other hand, many questions are left hanging.

First, can we turn the ideas about maximum and minimum useful integration times into a theory of optimal filtering that would predict, quantitatively, the form of the impulse responses in Fig 47? We should be able to do this, but I don't think anyone has really managed to get it right. There have been some serious attempts, but I think the issue still is open. One might also wonder whether it even makes sense to formulate this problem for individual bacteria, as opposed to looking at competition or cooperation in a population; this is related to the question of what, precisely, one thinks is being optimized by the behavior. It seems likely that any theory of optimal strategies will predict that this optimum is context dependent; here we should note that quantitative characterization of chemotactic behavior has not been pursued under a very wide range of stimulus conditions, so we may be missing the data we need to test

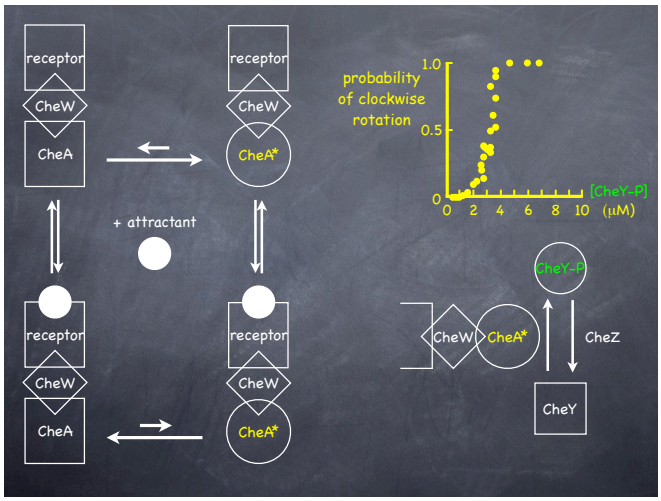

FIG. 48 Biochemical amplification in the chemotactic response. [Redraw this to make it more obvious that there is a cascade, as in rod photoreceptors.] At left, binding of chemoattractants to their receptors shifts the equilibrium between active and inactive forms of the kinase CheA. At right, the active kinase phosphorylates CheY, and this is balanced by the action of the phosphatase CheZ. CheY∼P binds to the flagellar motor and promotes clockwise rotation, which drives tumbling. The motor is extremely sensitive to small changes in the CheY∼P concentration; data redrawn from Cluzel et al (2001).

such theories when they emerge.

The second question is about the mechanisms that make possible the extreme sensitivity of chemotaxis. Much progress has been made, although again some issues are open. As with the rod cell, there is a cascade of biochemical events that leads from input (here, binding to receptors on the cell surface) to output (direction of motor rotation). Since input and output are spatialy separated, it is not surprising to find that there is an internal signaling molecule that diffuses through the cell. In rods, this is a small molecule (cGMP), but for bacterial chemotaxis it is a protein called CheY. More precisely, this protein can be phosophorylated, and in its phosphorylated form CheY∼P it binds to the motor and favors clockwise rotation. The receptor molecules on the cell surface are coupled almost directly to the kinase CheA that phosphorylates CheY, as shown schematically in Fig 48. Working backward from the output, we would like to know how the rotational bias of the motor depends on the concentration of CheY∼P.

To measure the bias vs CheY∼P, one has to do many tricks. It's relatively easy to measure the bias of the motor, either in experiments where the cell is tethered or where it is laying on a slide and one motor stump is sticking up with a bead attached. To know the concentration of a protein in a single cell, we need to make the protein visible, and so this is done by genetic engineering, replacing the normal CheY with a fusion between this protein and the green fluorescent protein [put clear discussion of GFP in the first place where it comes up—perhaps here?], and arranging for the expression of this fusion protein to be controlled by signals that can be applied externally. Finally, we need to know the concentration of the phosphorylated form of the protein, and this is very difficult. But once phosphate groups are attached to a protein, they stay there until removed by another enzyme (the phosphotase). So, if we genetically engineer the bacterium to remove the phosphotase, we will surely screw up the overall chemotactic response, but we can then be sure that all the CheY will be in its phosphorylated state. The result of all this is shown in Fig 48.

Problem 57: Absolute concentration measurements. In this problem you should try to understand how Cluzel et al were able to put the CheY∼Pconcentration on an absolute scale. Bacteria can be engineered to make a fluorescent version of many naturally occurring proteins. While the fluorescence signal that we then see under a microscope is proportional to the number of molecules under illumination, it can be difficult to measure the proportionality constant in an independent experiment. One can circumvent this problem by watching small numbers of molecules diffusing randomly in and out of an illuminated volume inside an individual cell and using the *variance* in the fluorescence intensity, along with its mean value, to make an absolute measurement of the concentration of the molecules.[43]

(a.) Explain (qualitatively) how this measurement might work. What do you gain by using both the variance and the mean of this signal? How can the fluctuating fluorescence signal be analyzed further to give an estimate of the protein diffusion constant?

(b.) Now let's convert the above intuition into a quantitative framework for analysis of the data. Consider the concentration $c(\vec{x},t)$ of fluorescent molecules at different points in space and time. It fluctuates and the deviation δc of the concentration from its average value $\bar{c}$ is uncorrelated between different points in space (but the same instant of time). Show that the analytic statement

$$\langle \delta c(\vec{x},t)\delta c(\vec{x}',t)\rangle = \bar{c}\delta(\vec{x}-\vec{x}') \tag{322}$$

of this fact is equivalent to the 'intuitive' remark that the variance of the number of molecules in a volume is equal to the mean number.

(c.) If the system starts with some fluctuation in the concentration $c(\vec{x},0) = \bar{c} + \delta c(\vec{x},0)$, this profile will relax according to the diffusion equation. Since the diffusion equation is linear, this means that the profile of fluctuations at time t, $\delta c(\vec{x},t)$, can be written as a linear operator acting on the initial condition $\delta c(\vec{x},0)$. Show that this linear relationship can be written as

$$\delta c(\vec{x},t) = \int d^3y \left(\frac{1}{\sqrt{4\pi Dt}}\right)^3 \exp(-|\vec{x}-\vec{y}|^2/4Dt)\ \delta c(\vec{y},0) \tag{323}$$

where D is the diffusion constant.

---

[43] Some of the ideas in this problem will, admittedly, be clearer after the discussion in the next section. Still, this should be workable now, and may provide a useful introduction to what comes next. This problem was originally designed as part of a general examination for Physics PhD students, written together with Curt Callan.

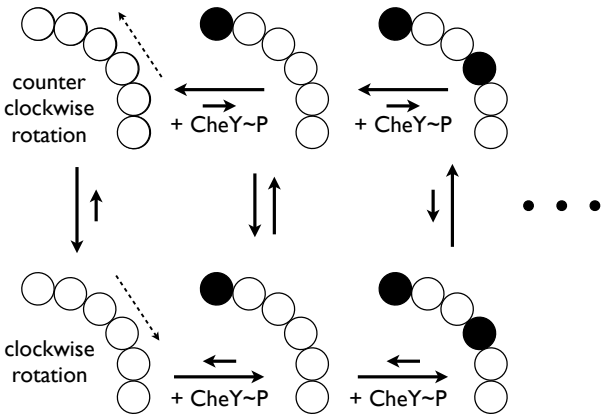

FIG. 49 A model for the modulation of rotor bias by binding of CheY∼P. CheY∼P molecules bind independently to multiple sites around a ring. When all sites are empty the equilibrium favors the counterclockwise rotating state. Binding is stronger to the clockwise state, however, so that as more sites are occupied the equilibrium shifts.

(d.) When we bring light to a focus under the microscope, we effectively weight the points around the focus with a Gaussian function, so that the light intensity collected from the fluorescent molecules will be proportional to

$$s(t) = \int d^3x\, c(\vec{x}, t) \exp(-|\vec{x}|^2/\ell^2) \tag{324}$$

where ℓ is the size of the focal region (roughly the size of the wavelength of light). Using the results above, show that the temporal correlation function of this signal is given by

$$\langle \delta s(t) \delta s(0) \rangle \propto (|t| + \tau)^{-3/2}, \tag{325}$$

and relate the correlation time τ to the diffusion constant D and the size of the focal region ℓ. As a hint, note that in doing the multidimensional Gaussian convolution integrals that show up in the last step of this computation, it is a good idea to do them Cartesian coordinate by Cartesian coordinate. This gives a precise method for extracting the diffusion constant from the fluctuating fluorescence signal.

What we see most clearly from Fig 48 is that the motor is remarkably sensitive to small changes in concentration of CheY∼P. One can fit a function of the form

$$P_{\rm cw} = \frac{c^n}{c^n + K^n}, \tag{326}$$

with $K \sim 3\,\mu$M and $n \sim 10$, although the data are almost within errors of being a step function. "Hill functions" of this form often are interpreted to mean that n molecules bind together and trigger the output that we are measuring; these and other ideas about the cooperative response of biological molecules are reviewed in Appendix A.4.

In this case it might make more sense to think about a model as in Fig 49, which is a version of the Monod–Wyman–Changeux model for cooperativity. Here we imagine multiple binding sites arrayed around a ring. CheY∼P molecules bind independently to each site, but the strength of the binding depends on whether the whole structure is rotating clockwise or counterclockwise. Qualitatively, if binding is stronger in the clockwise state, then increasing the concentration of CheY∼P will shift the equilibrium toward the clockwise state.

Quantitatively, we can work out the predictions of the model in Fig 49 using statistical mechanics, on the hypothesis that all the binding events and the structural transitions of the motor between clockwise and counterclockwise states come to equilibrium. One might worry about the latter assumption—after all, if the motor were truly at equilibrium it wouldn't be rotating and generating force—but let's proceed. Consider one possible state of the system, say clockwise rotation with m out of the n sites filled by CheY∼P molecules. We need to assign this state a weight in the Boltzmann distribution. We can assume that the clockwise state has an intrinsic (free) energy $E_{\rm cw}$. With k molecules bound, the energy is lowered by $mF_{\rm cw}$, where $F_{\rm cw}$ is the binding energy in the clockwise state, but we also had to take these k molecules out of solution, and this shifts the free energy by m times the chemical potential, $m\mu = mk_BT\ln(c/c_0)$, where c is the concentration of CheY∼P and c_0 is a reference concentration. Finally, since the m occupied sites could chosen out of the n possibilities in many ways, there is a combinatorial factor. Putting these terms together we have

$$\binom{n}{m} \exp\left[-\frac{1}{k_BT}\left(E_{\rm cw} - mF_{\rm cw} - mk_BT\ln(c/c_0)\right)\right] = \binom{n}{m}\left(\frac{c}{K_{\rm cw}}\right)^m e^{-E_{\rm cw}/k_BT},$$

where $K_{\rm cw} = c_0 e^{-F_{\rm cw}/k_BT}$. To compute the probability of being in the clockwise state we have to sum over all the different occupancies, and normalize by the partition function, which includes a sum over the counterclockwise states:

$$P_{\rm cw} = \frac{1}{Z}\sum_{m=0}^{n}\binom{n}{m}\left(\frac{c}{K_{\rm cw}}\right)^m e^{-E_{\rm cw}/k_BT} \tag{327}$$

$$= \frac{1}{Z} e^{-E_{\rm cw}/k_BT}(1 + c/K_{\rm cw})^n, \tag{328}$$

where

$$Z = e^{-E_{\rm cw}/k_BT}(1 + c/K_{\rm cw})^n + e^{-E_{\rm ccw}/k_BT}(1 + c/K_{\rm ccw})^n. \tag{329}$$

We can put this result in a more compact form,

$$P_{\rm cw} = \frac{1}{1 + \exp\left[\theta - g(c)\right]} \tag{330}$$

$$\theta = (E_{\rm cw} - E_{\rm ccw})/k_BT \tag{331}$$

$$g(c) = n\ln\left(\frac{1 + c/K_{\rm cw}}{1 + c/K_{\rm ccw}}\right). \tag{332}$$

Notice that if $K_{\rm cw} \ll c \ll K_{\rm ccw}$, then this becomes the Hill function in Eq (326).

Problem 58: MWC model of rotor bias. Explore the parameter space of the model we have just described. Are there regimes, other than $K_{cw} \ll c \ll K_{ccw}$, where one can reproduce the steep dependence of P_{cw} on c observed by Cluzel et al (2001)? Keep in mind that the actual number of binding sites n could be very large.

So part of the answer to how the the bacterium is so sensitive to small changes in the external concentration of attractants or repellents is that the motor is very sensitive to small changes in the concentration of CheY∼P. This is not implausible, since the structure of the motor (which is complicated) suggests locations for as many as $n = 34$ sites where CheY∼P could bind around a ring of radius $R \sim 45\,\text{nm}$.

Having such strong sensitivity to the CheY∼P concentration means that, in roughly the one second it takes for the motor to switch once, one can be sure whether the concentration was $\delta c/c \sim 1/n \sim 10\%$ above or below the critical value $c = K$. But from Berg and Purcell we might expect that there is a limit on this precision set by random arrival of the CheY∼P molecules at the motor, and this should be $\delta c/c \sim 1/\sqrt{DRc\tau_{\text{avg}}}$, treating the whole motor ring as one big receptor. With diffusion constants for proteins, including CheY, in the range of $D \sim 1\mu\text{m}^2/\text{s}$, this suggests that the limit with one second of integration is not much smaller than 10% (see more details in the next lecture). So, cooperative action of many signaling molecules generates a steep slope, but the system still has to suppress other sources of noise since even this last step in the cascade of events is operating close to the fundamental limits set by noise considerations.

The observations on the sensitivity of the motor tell us that the bacterium can generate a significant response even from a small fractional change in the concentration of CheY∼P. Still, we need to understand the biochemical processes that lead from essentially single molecular events to these quasi–macroscopic changes in molecule number.[44] [Probably want to say a few words about the sources of gain: activity of CheA*, and the cooperativity among receptors that allows one ligand to activate many CheAs. Need to learn more about the numbers here. Might be nice to compare MWC–style model of motor with MWC–style model of receptors. At the end of the day, is this similar to the rod cell or not? Can we conclude that we understand the gain?]

[44] At $c \sim 3\,\mu\text{M}$, a cell with volume $\sim 1\,\mu\text{m}^3$ has ~ 2000 molecules of CheY∼P, so even a ten percent change in concentration involves hundreds of molecules.

Even if we consider the origins of gain to be understood, there is a major problem. Figure 48 shows that extreme sensitivity must coexist with a very tight regulation, since if the concentration of CheY∼P drifts far away from $c \sim K$, the cell loses all sensitivity to changes. This combination of sensitivity to small changes without accumulation of large variations poses significant problems, which we will take up in the next Chapter.

The last of the major questions left open by the Berg–Purcell analysis is whether we do a full, honest calculation that leads to the their limit on the precision of concentration sensing? What Berg and Purcell wrote down makes absolutely no reference to the messy details of what actually happens to molecules as they are counted. This could be wonderful, because it would mean that can say something about the limits to precision in *all* biochemical signaling systems, regardless of details. Alternatively, the absence of details might be a disaster, a clue that we have simply missed the point.

As mentioned at the start of this section, chemical signaling in ubiquitous in biological systems, and chemotaxis provided us with one clear example where we could think about the limits to counting molecules. We would like to know if these limits can be made rigorous, and

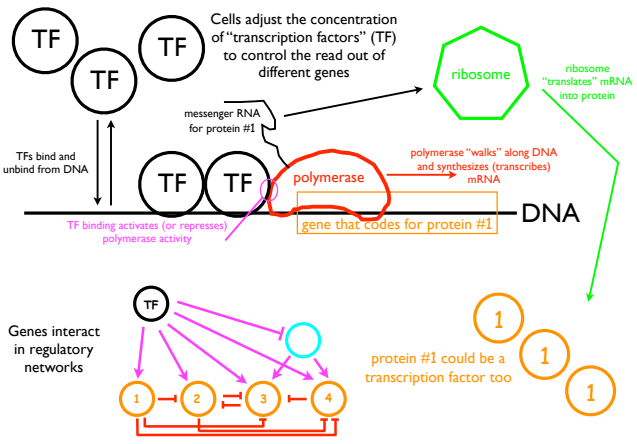

FIG. 50 Control of gene expression by transcription factors. Synthesis of a protein involves transcription of the DNA coding for that particular protein, and translation of the resulting mRNA. An important component of control in these systems is the binding of transcription factors to the DNA, at specific sites near the start of transcription, in the promoter or enhancer region. Transcription factors are themselves proteins, so this regulatory process naturally leads to a network of interactions; here we focus, for simplicity, on one input (the concentration of the transcription factor) and one output (the concentration of protein #1). Note that in bacteria all of this happens in one compartment, while in eukaryotic cells the DNA is in the nucleus and mRNA is transported out to the cytoplasm, where translation occurs. Nothing in this figure is to scale. [redraw figure to get rid of the network, which here is a distraction]

if they can be applied to processes that occur inside of cells, rather than just to the sensing of external signals as in chemotaxis. To see what is at stake, let's think about the regulation of gene expression (Fig 50). We recall that every cell in our bodies has the same DNA. What makes a liver cell different from a neuron in your brain is that it reads out or "expresses" different genes, making different proteins. Importantly, this is not just a discrete choice made once in your lifetime. Given that certain proteins are being made, the numbers of these molecules are constantly adjusted to match the needs of the cell. This happens also in bacteria, which adjust, for example, the concentrations of the enzymes needed to metabolize different nutrients that might or might not be present in the environment; much of what we know about the regulation of gene expression has its roots in work on this sort of metabolic control in bacteria.

There are many ways in which gene expression is controlled. As a simple example, note that if we want to regulate the number of proteins in the cell we can change either the rate at which they are made or the rate at which they are degraded, and both of these things happen. The synthesis of a protein involves two very different steps, transcription from DNA to messenger RNA and translation from mRNA to protein, and again there is regulation of both processes. All this being said, we will focus our attention on the regulation of transcription, that is the reading of the DNA template to make mRNA.[45]

In order to make mRNA, a complex of proteins (including the RNA polymerase) must bind to the DNA and 'walk' along it, spewing out the mRNA polymer as it walks. In order for all of this to happen, the RNA polymerase has to find the right starting point. One can imagine that this can be inhibited simply by having other proteins bind to nearby sites along the DNA. Alternatively, binding of proteins to slightly different positions near the starting point could help the RNA polymerase to find its way. Both of these things happen: proteins called transcription factors can act both as repressors and as activators of mRNA synthesis. The key step in this regulation is thought to be the binding of the transcription factors to specific sites near the RNA polymerase start site, as schematized in Fig 50; the whole segment of DNA involved in the control and initiation of transcription is called the "promoter." In higher organisms, the regions involved in regulation can be very large indeed, and usually are called "enhancers" to avoid conjuring the simplified image in Fig 50, which is more literally applicable in bacteria. Binding sites are specific because the transcription factor protein is selective for particular DNA sequences, and much can be said about the nature of this specificity. For now the important point is that such regulatory systems are, in effect, sensors of the transcription factor concentration.

Problem 59: Autoregulation. Perhaps the simplest model of transcriptional regulation is one in which a gene regulates its own expression. Let the concentration (or number of molecules) of the protein be g, and assume that n of these molecules bind cooperatively to the promoter region of the gene. If the binding activates expression, and proteins are degraded in a simple first order process with lifetime τ, then it is plausible that the dynamics of g are given by

$$\frac{dg}{dt} = r_{\max}\frac{g^n}{g^n + g_{1/2}^n} - \frac{g}{\tau}. \tag{333}$$

(a.) Explain the significance of the parameters $r_{\max}$, n and $g_{1/2}$. Show that there is a range of these parameters in which the system is bistable. More precisely, show that you can find three steady states, and that two of these are stable and one is unstable. What are the time constants for relaxation to these steady states? How do these times compare with the lifetime τ of the protein?

(b.) Really the protein binding regulates the synthesis of mRNA, which in turn is translated by the ribosomes into protein. If m is the mRNA concentration (or number of molecules), then a plausible set of equations is

$$\frac{dm}{dt} = e_{\max}\frac{g^n}{g^n + g_{1/2}^n} - \frac{m}{\tau_{\rm m}} \tag{334}$$

$$\frac{dg}{dt} = r_{\rm trans}m - \frac{g}{\tau_{\rm p}}, \tag{335}$$

where $e_{\max}$ is the maximal transcription ("expression") rate, $r_{\rm trans}$ is the rate at which mRNA molecules are translated into protein, and the lifetimes of protein and mRNA are $\tau_{\rm p}$ and $\tau_{\rm m}$, respectively. Under what conditions will this more complete model be well approximated by the simpler model above? Are the steady states of the two models actually different? What about their stability?

(c.) Suppose that instead of activating its own expression, the protein acts as a repressor of its own expression. Find the analog of Eq (333) in this case and show that there is only one steady state, and that this state is stable.

(d.) Expand your discussion of the auto–repressor to include the mRNA concentration, as in Eq's (334, 335). Find the steady state and linearize the equations around this point. Do you find exponential relaxation toward the steady state for all values of the parameters? Is it possible for the steady state to become unstable? Explain qualitatively what is happening, and go as far as you can in analyzing the situation analytically.

The binding sites along DNA for the transcription factors have linear dimensions measured in nanometers, perhaps $a \sim 3\,$nm. The diffusion constants of proteins in the interior of cells is in the range of $D \sim 1\,\mu{\rm m}^2/{\rm s}$. Many transcription factors act at nanoMolar concentrations, and it is useful to note that $1\,{\rm nM} = 0.6\,{\rm molecules}/\mu{\rm m}^3$. Putting these together we have $Dac \sim 1.8\times 10^{-3}\,{\rm s}^{-1}$. Thus, the Berg–Purcell limit predicts that the smallest

[45] For a bit about the basics of DNA structure, see Appendix A.5.

changes in transcription factor that can be reliably detected are

$$\frac{\delta c}{c} \sim \frac{1}{\sqrt{Dac\tau_{\rm avg}}} \sim \sqrt{\frac{10\,{\rm min}}{\tau_{\rm avg}}}. \qquad (336)$$

Taken at face value, this suggests that truly quantitative responses—say, to 10% changes in transcription factor concentration—would require hours of integration. This is seldom plausible.

One should not take this rough estimate too literally. I think the message is not the exact value of the limiting precision, but rather that once concentrations fall to the nM range, small changes will be very hard to detect. If cells do detect these small changes, then almost certainly they will be bumping up against the physical limits set by counting molecules, assuming that Berg and Purcell give us a good estimate of these limits. So, this is what we need to check.

In Appendix A.6, we look in detail at how to make the Berg–Purcell limit more rigorous. The key idea is that fluctuations in concentration, and in many examples of binding to receptor sites, represent fluctuations in thermal equilibrium, and thus are susceptible to the same analyses as Brownian motion, Johnson noise, and other examples of thermal noise. These analyses show how one can separate the limiting noise level from the extra noise that is associated with all the biochemical complexities which Berg and Purcell ignored. The result, then is that the Berg–Purcell argument can be made rigorous, both for single receptors and for arrays of receptors, and their simple formula gives us a *lower bound* on the noise in biochemical signaling. This is important because, as noted at the start of this discussion, the Berg–Purcell limit doesn't make reference to any of the detailed biochemistry of what happens when the signaling molecules bind to their targets. Rather, the limit depends on the physical nature of the signal itself. The fact that we can make the Berg–Purcell argument rigorous encourages us to look more broadly and see if there are other cases in which biological systems approach these physical limits to their signaling performance.

Would like to discuss chemotaxis in larger cells—neutrophils, Dictyostelium, ...

Another important example of chemotaxis occurs during the development of the brain. Individual neurons start as relatively compact cells, and then extend their axons to find the other cells with which they must make synapses. This processes is guided by gradients in a variety of signaling molecules. Although there are many beautiful observations on these phenomena in vivo, it is not so easy to do a controlled experiment where one allows cells to migrate in well defined gradients. One approach to this is shown in Fig [reproduce figures from Rosoff et al], where cells grow in a collagen matrix that is "printed" with droplets of growth factor at varying densities. Relatively quickly, diffusion acts to smear the rows of drops into a continuous gradient, which can be directly observed when the molecules are labelled with fluorophores. These measurements also allow an inference of the diffusion constant in this medium, $D \sim 8 \times 10^{-7}\,{\rm cm}^2/{\rm s}$. The growth cones which guide the axon have linear dimensions $a \sim 10\,\mu{\rm m}$, and these experiments found that sensitivity to gradients is actually maximal in a concentration range near $c \sim 1\,{\rm nM}$. Under these conditions, then, we have $Dac \sim 500\,{\rm s}^{-1}$. Quite astonishingly, however, the cells seem to grow differentially in the direction of gradients that correspond to concentration differences of order one part in one thousand across the diameter of the growth cone. In order for this signal to be above the Berg–Purcell limit on the noise level, the cell must integrate for $\tau_{\rm avg} \sim 2000\,{\rm s}$, a reasonable fraction of an hour.

In truth, we don't know the time scale over which growth cones are integrating as they decide which way to turn, even in the more controlled in vitro experiments. We do know that the pace of neural development is slow—hours to days rather than minutes. Qualitative aspects of axonal behavior are consistent with the idea that the time scales of their movements are determined by the need to integrate long enough to generate reliable directional signals, from the rapid "exploration" by cellular appendages to the dramatic slowing down near critical decision points, such as the optic chiasm where the axons of ganglion cells emerging from the retina must decide whether to go toward the right or left half of the brain.[46] It is attractive to think that the reliability with which cells in our brain find their targets is set by such basic physical principles, but we don't quite have enough data to say this with certainty.

Let us return to the problem that motivated our search for generality, the transcriptional regulation of gene expression. Until the last decade, there were essentially no direct measurements on the reliability of such regulatory mechanisms. Before we look at the new data, though, we need one more set of theoretical ideas.

Proteins are synthesized and degraded, and the simplest assumption is that these are single kinetic steps. Suppose we start just with synthesis, at some rate s molecules per second. We have seen that rate constants should be interpreted as the probability per unit time for individual molecular events. Thus, if we ask about the probability of finding exactly N molecules in the system at time t, this probability $P(N;t)$ obeys the "master

[46] At these decision points it seems likely that the cells must reach rather high signal–to–noise ratios, since the error probabilities are small. [can we say something quantitative here?]

equation"

$$\frac{\partial P(N;t)}{\partial t} = sP(N-1;t) - sP(N;t), \tag{337}$$

except of course at $N = 0$ where we have

$$\frac{\partial P(0;t)}{\partial t} = -sP(0;t). \tag{338}$$

We can solve these equations iteratively. We start with no molecules, so $P(0,0) = 1$, while $P(N \neq 0, 0) = 0$. Then Eq (338) tells us that

$$P(0,t) = e^{-st}. \tag{339}$$

If we substitute into Eq (337) for $P(1,t)$, we have

$$\frac{\partial P(1;t)}{\partial t} = -sP(1;t) + sP(0;t) \tag{340}$$

$$\Rightarrow P(1,t) = \int_0^t dt' e^{-s(t-t')} sP(0;t) \tag{341}$$

$$= \int_0^t dt' e^{-s(t-t')} s e^{-st} \tag{342}$$

$$= se^{-st} \int_0^t dt' = e^{-st}(st). \tag{343}$$

We can go through the same calculation for $P(2;t)$:

$$P(2;t) = \int_0^t dt' e^{-s(t-t')} sP(1;t') \tag{344}$$

$$= e^{-st} \int_0^t dt' s^2 t' \tag{345}$$

$$= e^{-st} \frac{(st)^2}{2}. \tag{346}$$

This suggests that, for all N,

$$P(N;t) = e^{-st} \frac{(st)^N}{N!} \tag{347}$$

Problem 60: Checking the Poisson solution. Verify that Eq (347) solves the master equation describing a single synthesis reaction at rate s, Eq (337).

Equation (347) is telling us that, as the synthesis reaction proceeds, the number of molecules that has been synthesized obeys the Poisson distribution. From what we have said about the Poisson distribution in the discussion of photon counting (Section I.A and Appendix A.1), you should recognize that the mean number of molecules is

$$\langle N \rangle \equiv \sum_{N=0}^{\infty} NP(N;t) = st, \tag{348}$$

which makes perfect sense. Further, the variance in the number of molecules is equal to the mean, at all times.

[This discussion is written without any figures. Maybe we need some schematics?]

What happens when we add degradation to this picture? Now the state of the system can change in several ways, all of which will modify the probability that there are exactly N molecules. First, synthesis can cause the N molecules to become $N+1$, reducing $P(N,t)$. Second, we can have the transition from $N-1$ to N molecules, which increases $P(N,t)$. Note that these first two terms were already present in our simpler model. The third process is where degradation takes N molecules and eliminates one, resulting in $N-1$ molecules. Since each molecule makes its transitions independently, the rate of this process must be proportional to N, and this reduces $P(N,t)$. Finally, if there were $N+1$ molecules, degradation results in N, increasing $P(N,t)$; again because each molecule is independent, the rate of this process must be proportional to $N+1$. Putting the terms together we have

$$\frac{\partial P(N;t)}{\partial t} = -sP(N;t) + sP(N-1;t) - kNP(N;t) + k(N+1)P(N+1;t), \tag{349}$$

where k is the probability per unit time for the decay of one molecule.

Now it is possible for the synthesis and degradation reactions to balance, generating a steady state. In this steady state the distribution of the number of molecules must obey

$$0 = sP(N-1) - (s+kN)P(N) + k(N+1)P(N+1). \tag{350}$$

To solve this equation it is useful to regroup the terms,

$$-sP(N-1) + kNP(N) = -sP(N) + k(N+1)P(N+1). \tag{351}$$

where the left hand side now refers to the forward and backward rates between states with $N-1$ and N molecules, while the right hand side refers to the transitions between N and $N+1$. All that we require is that the

two sides be equal, but suppose we try to set each side separately to zero, which corresponds to "detailed balance" among the transitions into and out of each state. Then from the left hand side we have

$$\frac{P(N)}{P(N-1)} = \frac{s}{kN}, \tag{352}$$

while from the right we have

$$\frac{P(N+1)}{P(N)} = \frac{s}{k(N+1)}. \tag{353}$$

But except for $N \to N+1$, these are the same equation. Thus, the steady state of this system does obey detailed balance, and we can solve by iterating Eq (352):

$$P(1) = \frac{s}{k}P(0) \tag{354}$$

$$P(2) = \frac{s}{2k}P(1) = \frac{(s/k)^2}{2}P(0) \tag{355}$$

$$P(3) = \frac{s}{3k}P(2) = \frac{(s/k)^3}{3!}P(0), \tag{356}$$

and, in general,

$$P(N) = \frac{(s/k)^N}{N!}P(0). \tag{357}$$

Finally we can fix the value of $P(0)$ by insisting that the distribution be normalized, and we find

$$P(N) = e^{-M}\frac{M^N}{N!}, \tag{358}$$

which again is the Poisson distribution, with mean $M = s/k$.

Problem 61: The diffusion approximation. If N is not too small we expect that $P(N;t)$ and $P(N \pm 1;t)$ are not too different. Thus we should be able to approximate using a Taylor series,

$$P(N \pm 1;t) \approx P(N;t) \pm \frac{\partial P(N;t)}{\partial N} + \frac{1}{2}\frac{\partial^2 P(N;t)}{\partial N^2}. \tag{359}$$

(a.) Show that this approximation turns the master equation in Eq (349) into something that looks more like the diffusion equation. What is the effective potential in which the "coordinate" N is diffusing?

(b.) Why does it make sense to stop the Taylor series after two derivatives? What happens if we stop after one?

(c.) How does the steady state solution that you obtain in the diffusion approximation compare with the exact solution (the Poisson distribution)?

Problem 62: Langevin equations for chemical kinetics. We know, as reviewed in Section II.A, that we can describe Brownian motion by either a diffusion equation or a Langevin equation. In more detail, we started with kinetics that, in the macroscopic limit, correspond to the dynamics

$$\frac{dN(t)}{dt} = s - kN(t). \tag{360}$$

We would like to describe the noisy version of these dynamics as

$$\frac{dN(t)}{dt} = s - kN(t) + \zeta(t), \tag{361}$$

where—inspired by the Brownian motion example—we expect that the noise $\zeta(t)$ is white, but the strength might depend on the state of the system, so that

$$\langle \zeta(t)\zeta(t')\rangle = T_{\rm eff}[N(t)]\delta(t-t'), \tag{362}$$

where to remind us of the analogy to Brownian motion we can refer to the noise strength as an effective temperature $T_{\rm eff}$.

(a.) Find the effective temperature that will reproduce the diffusion equation that you derived in the preceding problem.

(b.) If we integrate Eq (361) over a very small time interval $\Delta\tau$, we obtain

$$\Delta N \equiv N(t+\Delta\tau) - N(t) \tag{363}$$

$$= [s - kN(t)]\,\Delta\tau + \int_0^{\Delta\tau} dt'\zeta(t+t'). \tag{364}$$

But if $\Delta\tau$ is small enough, we know that the changes in the number of molecules should be $\Delta N = 0$ or $\Delta N = \pm 1$. Going back to the master equation [Eq 349], identify these transition probabilities. From these probabilities, show that the mean change in the number of molecules is the first term in Eq (364), $\langle \Delta N\rangle = [s - kN(t)]\Delta\tau$. Continuing, show that the variance in ΔN is given by $\langle(\delta\Delta N)^2\rangle = [s + kN(t)]\Delta\tau$.

(c.) To reproduce the variance in ΔN, we must have

$$\left\langle \left(\int_0^{\Delta\tau} dt'\zeta(t+t')\right)^2 \right\rangle = [s + kN(t)]\Delta\tau. \tag{365}$$

Use this, together with Eq (362), to show that

$$T_{\rm eff}[N(t)] = s + kN(t). \tag{366}$$

Does this agree with your result in (a.)?

So, these simplest of kinetic schemes for the synthesis and degradation of molecules predict that the distribution of the number of molecules ("copy numbers") should be Poisson. Certainly we can imagine kinetic schemes for which the fluctuations in copy number will be larger than Poisson. For example, if the simple picture of synthesis and degradation were correct for messenger RNA, but each mRNA leads to the synthesis of b proteins, then the mean number of proteins will be larger than the mean number of mRNA molecules by this factor b, $\langle N_{\rm p}\rangle = b\langle N_{\rm mRNA}\rangle$, but the variance will be larger by a factor of b^2, $\langle(\delta N_{\rm p})^2\rangle = b\langle(\delta N_{\rm mRNA})^2\rangle$. Thus, if we count protein molecules, the variance will be larger than the mean, $\langle(\delta N_{\rm p})^2\rangle = b\langle N_{\rm p}\rangle$, and hence the protein copy numbers are more variable than expected from the Poisson distribution. Notice that this is true even though we have assumed that the translation from mRNA to protein is completely noiseless, with each mRNA making exactly b proteins. Variance beyond the Poisson expectation here arises simply from amplification. This is exactly the same argument made about photons and spikes from ganglion cells in the retina, in Section I.D.

With this background, what can we measure? Counting protein molecules is not easy. Over the last decades,

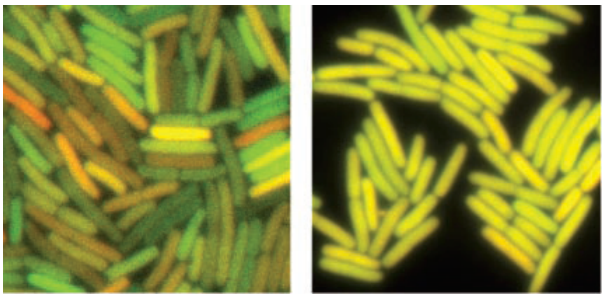

FIG. 51 Noise in the regulation of gene expression, from Elowitz et al (2002). A population of *E coli* express two fluorescent proteins of different colors, CFP and YFP, both under the control of the *lac* repressor. At left, expression is repressed, copy numbers are low, and color variations are substantial. Thus, although the two genes see the same regulatory signals, there is intrinsic variation in the output. At right, repression is relieved, expression levels are higher, and color variations are substantially smaller.

we have seen a huge improvement in the methods of optical microscopy, to the point where we can literally see the light emitted from a single fluorescent molecule. But most biological molecules, and most proteins in particular, are not fluorescent. Indeed, until relatively recently the only proteins with interesting spectroscopic signatures in the visible part of the spectrum (e.g., the visual pigments and the heme proteins) involved a smaller molecular cofactor bound to the protein (retinal, heme). These cofactors are synthesized by separate, often complex pathways. Thus while it might be possible to engineer a cell to make a pigment protein just by splicing the relevant gene into its genome, it would be almost impossible to introduce the entire synthetic machinery for the cofactor. This is why the discovery of the green fluorescent protein in a species of jellyfish turned out to be so important. In contrast to the proteins which require cofactors for their fluorescence, these molecules are intrinsically fluorescent [Need a figure showing structure, point to why this is possible, etc.. Maybe this discussion should come earlier?] Since the isolation of the original GFP, many variants have been synthesized, in a variety of colors.

The simplest experiment to probe noise in the expression of a gene would be to introduce the gene for GFP into a bacterium, and just look at the levels of fluorescence—the brightness will be proportional to the number of molecules, and with luck we can even calibrate the proportionality factor. But expression levels could vary for uninteresting reasons. Cells vary in size as they grow and divide. There can be variations in the number of ribosomes, which will change the efficiency of translation but it probably doesn't make sense to call these variations "noise." How do we separate all these different sources of variation from genuine stochasticity in the processes of transcription and translation?

If we go back to Fig 50, we see that the transcription of a gene into RNA is controlled by the binding of transcription factor proteins to a segment of DNA called the promoter or (in higher organisms) enhancer region. Suppose that we make two copies of the same promoter, put one next to the gene for a green fluorescent protein and one next to the gene for a red fluorescent protein, and then reinsert both of these into the genome. Now all variations in the state of the cell that affect the overall efficiency of transcription and translation will change the levels of green and red proteins equally. If the regulatory signals were noiseless, and the independent processes of transcription and translation of the two proteins were similarly deterministic, then every cell would be perfectly yellow, having made equal amounts of green and red protein; cells might differ in their total brightness, but the balance of red and green would be perfect. On the other hand, if there really is noise in transcription and translation, or their regulation, then the balance of red and green will be imperfect, and if we look a population of genetically identical cells they will vary in color as well as in brightness.

Figure 51 shows that our qualitative expectations for a "two color" experiment are borne out in real experiments on *E coli*, although "red" and "green" are actually yellow and cyan. In this experiment, the two fluorescent proteins are under the control of the *lac* promoter. In the native bacterium, this promoter controls the expression of enzymes needed for the metabolism of lactose, and if there is a better source of carbon available (or if lactose itself is absent) the bacteria don't want to make these enzymes. There is a transcription factor protein called lac repressor which binds to the *lac* promoter and blocks transcription. By changing environmental conditions, one can tap into the signals that normally tell the bacterium that it is time to turn on the *lac*–related enzymes, and turn off the repression by inactivating the repressor proteins. Thus, not only can we get *E coli* to make two colors of fluorescent protein, we can even arrange things so that we have control over the mean number of proteins that will be made. Everything that we have said thus far about noise in synthesis and degradation reactions predicts that if the cell makes more protein on average, then the fractional variance in how much protein is made should be reduced, and this is exactly what we see in Fig 51.

More quantitatively, in Fig 52 we see the decomposition of the variations into an "extrinsic" part that changes the two colors equally and an "intrinsic" part that corresponds to relative variations in the expression

of the two proteins that are under nominally identical control. If synthesis and degradation of proteins were a Poisson process, then we expect from above that the variance would be equal to the mean; amplification of Poisson fluctuations in mRNA count would leave the variance proportional to the mean. Even if the Poisson model is exact, if we can't calibrate the fluorescence intensity to literally count the molecules, again all we could say the that the variance of what we measure will be proportional to the mean. In fact, the data are described well by

$$\frac{\langle (\delta F)^2 \rangle}{\langle F \rangle} = \frac{A}{\langle F \rangle} + B, \tag{367}$$

where the fluorescence is normalized so that the mean under conditions of maximal expression is one, and $A = 7 \times 10^{-4}$ and $B = 3 \times 10^{-3}$. If $B \to 0$, this is exactly the prediction of the Poisson model, and indeed B is small. Importantly, we can see the decrease in the fractional noise level with the increase in the mean. The absolute numbers also are interesting, since they tell us that cells can—at least under some conditions—set the expression level of a protein to an accuracy of better than 10%.

It has been appreciated for decades that the initial steps in the development of embryos provides an excellent laboratory in which to study the regulation of gene expression. As we have mentioned several times, what makes the different cells in our body different is, fundamentally, that they express different proteins. These differences in expression have a multitude of consequences, but the first step in making a cell commit to being one type or another is to turn on (and off) the expression of the correct set of genes. At the start, an embryo is just one cell, and through the first several rounds of cell division it is plausible that the daughter cells remain identical. At some point, however, differences arise, and these are the first steps on the path to differentiation, or specialization of the cells for different tasks in the adult organism.

A much studied example of embryonic development is the fruit fly *Drosophila melanogaster*. We will learn much more about this system in Section III.C, but for now the key point is that in making the egg, the mother sets the initial conditions for development in part by placing the mRNA for key proteins—referred to as the "primary morphogens"—at cardinal points in the embryo. As these messages are translated, the resulting proteins diffuse through the embryo, and act as transcription factors, activating the expression of other genes. An example is Bicoid, for which the mRNA is localized at the (eventual) head; the diffusion and degradation of the Bicoid (Bcd) protein leads to a spatial gradient in its concentration, and we can visualize this by fixing and stain-

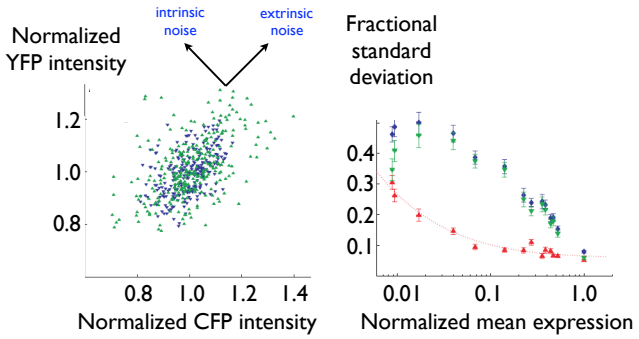

FIG. 52 Separating intrinsic and extrinsic noise, from Elowitz et al (2002). At left, a scatter plot of the fluorescence from the two different proteins show the decomposition into variations in the overall efficiency of transcription and translation ("extrinsic" noise) and fluctuations that change the two expression levels independently ("intrinsic" noise). At right, while the total variance has no simple dependence on the mean expression level, the intrinsic noise goes down systematically as the mean expression level goes up. Quantitatively, we plot the standard deviation σ in fluorescence level, divided by the mean m, as a function of the mean. The dotted line is from Eq (367).

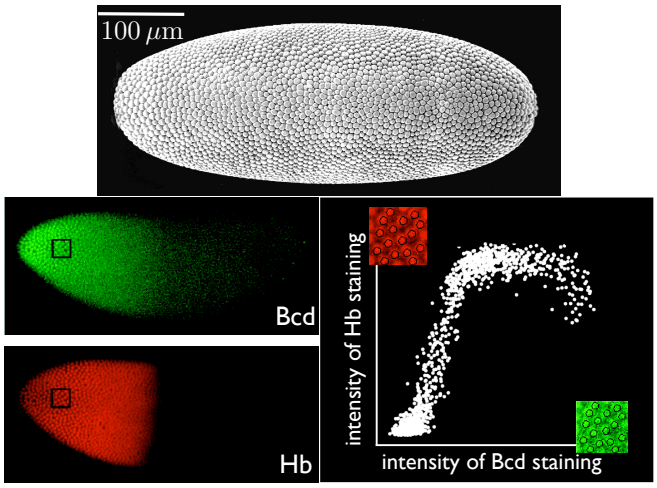

FIG. 53 Bicoid (Bcd) and Hunchback (Hb) in the early *Drosophila* embryo. At top, an electron micrograph of the embryo in cell cycle fourteen, with thousands of cells in a single layer at the surface (image courtesy of EF Wieschaus). At the bottom left, the embryo has been exposed to antibodies against the proteins Bcd and Hb, and these antibodies in turn have been labelled by green and red fluorophores, respectively; the fluorescence intensity should be proportional to the protein concentration, perhaps with some background. Bicoid is a transcription factor that activates the expression of Hunchback, and at the bottom right we see a scatter plot of the output (Hb) vs input (Bcd), where each point represents the state of one nucleus from the images at left; from Gregor et al (2007b).

ing the embryo with fluorescent antibodies, as shown in Fig 53. A more modern approach is to fuse the gene for Bcd with a fluorescent protein and substitute this for the original gene; if one can verify that the fusion protein replaces the function of the original, quantitatively, then we can measure the spatial profile of Bcd in a live embryo. Among other things, this approach makes it possible to demonstrate that the fluorescence signal from antibody staining really is proportional to the protein concentration, so we can interpret the data from images such as those in Fig 53 quantitatively.

From our point of view, in constructing the embryo, the mother has created an ideal experimental chamber. After just a few hours, there are thousands of cells in a controlled environment, exposed to a range of input transcription factor concentrations that we can literally read out along the embryo. We can also measure the response to these inputs, for example the expression of the protein Hunchback shown in Fig 53. In fact the targets of Bcd are themselves transcription factors, so conveniently they localize back to the nucleus, and hence each nucleus gives us one data point for characterizing the input/output relation. Taking seriously the linearity of antibody staining we can plot the input/output relation between Bcd and Hb in appropriately normalized coordinates, as in Fig 54, and we can measure the noise in expression by computing the variance across the many nuclei that experience essentially the same input Bcd level.

The first thing we see from Fig 54 is that, consistent with the results from bacteria in Fig 52, the embryo can regulate the expression of Hunchback to $\sim 10\%$ accuracy or better across much of the relevant dynamic range. How does this compare with the physical limits? To measure the reliability of Hunchback's response to Bicoid, we should refer the noise in expression back to the input—if we want to change the output by an amount that is equal to one standard deviation in the noise, how much do we have to change the input? The answer is given by propagating the variance backwards through the input/output relation,

$$\langle(\delta \mathrm{Hb})^2\rangle = \left|\frac{d\langle \mathrm{Hb}\rangle}{d\ln c}\right|^2 \left(\frac{\delta c}{c}\right)^2_{\mathrm{eff}}, \tag{368}$$

where c is the concentration of Bcd, and $(\delta c/c)_{\mathrm{eff}}$ defined in this way should be comparable to the Berg–Purcell limit. In Fig 54 we see that this effective noise level drops down to $(\delta c/c)_{\mathrm{eff}} \sim 0.1$, so the system seems able to respond reliable to $\sim 10\%$ differences in concentration of the input transcription factor.

We have seen, in Eq (336) and the surrounding discussion, that responding reliably to 10% differences in transcription factor concentrations would be very difficult to detect, requiring hours of integration to push the noise level down to manageable levels. This seems generally implausible, but in the fly embryo it is impossible,

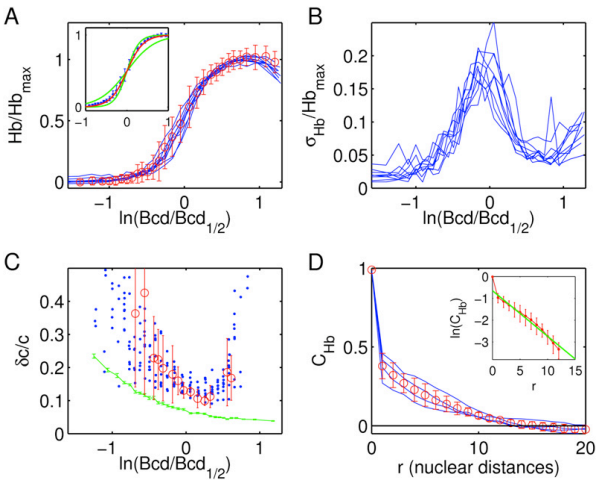

FIG. 54 Input/output and noise in the transformation from Bcd to Hb, from Gregor et al (2007b). (A) The input/output relation can be obtained starting from the scatter plot in Fig 53, normalizing the fluorescence intensities as relative concentrations, and then averaging the output Hb expression level across all nuclei that have essentially the same input Bcd level. Blue curves show results for several individual embryos, and red circles with error bars show the mean and standard deviation of Hb expression level vs Bcd input for a single embryo. The inset shows that these data are well fit by a Hill relation [see the discussion around Eq (326)] with $n = 5$ (in red), and substantially less well fit by $n = 3$ or $n = 7$ (in green). (B) The standard deviation of Hb output, measured across the multiple nuclei with the same Bcd input in single embryos; different curves correspond to different individual embryos. (C) Combining the input/output relation and noise levels, we obtain the effective noise level referred to the input, as in Eq (368); blue points are raw data, green line is an estimate of measurement noise, and red circles are the results of subtracting the measurement noise variance, with error bars computed across nine embryos. (D) Correlations in Hb expression noise in different nuclei, as a function of distance.

since the whole process from laying the egg to the establishment of the basic body plan (several steps beyond the expression of Hunchback) is complete within three hours or less. This apparent paradox depends on estimating some key parameters, but in the Bcd/Hb system these can be measured, and the solution to the problem does not seem to lie here.

Problem 63: Effective diffusion constants. Add a problem about the renormalization of diffusion constants by transient binding ... connect to noise levels, in a somewhat open ended second part.

On the other hand, the fly embryo is unusual in that, for much of its early development there are no walls between the cells. Thus, Hunchback mRNA synthesized

in one nucleus will be exported to the neighboring cytoplasm, and the translated protein should be free to diffuse to other nuclei. Thus the Hunchback level in one nucleus should reflect an average over the Bcd signals from many cells in the neighborhood. If Hb has a diffusion constant similar to that of Bcd, then in a few minutes the molecules can cover a region which includes ~ 50 nuclei, and averaging over 50 independent Bcd signals is enough to convert the required integration time from hours to minutes. If this scenario is correct, there should be correlations among the Hb expression noise in nearby nuclei, and this is what we see in Fig 54D. Indeed, the correlation length of the fluctuations is just what we need in order to span the minutes/hours discrepancy. These results suggest strongly that the reliability of the Hunchback response to Bicoid is barely consistent with the physical limits, but only because of spatial averaging.

Can we give a fuller analysis of noise in the Bcd/Hb system? In particular, we see from Fig 54B that the noise level has a very characteristic dependence on the input concentration, which we can also replot vs the mean output, as in Fig 55. This is an interesting way to look at the data, because in the limit where the Poisson noise of synthesis and degradation is dominant we should have

$$\langle(\delta\mathrm{Hb})^2\rangle_{\mathrm{Poisson}} = \alpha\langle\mathrm{Hb}\rangle, \tag{369}$$

where the constant α depends on the units in which we measure expression, but reflects the absolute number of independent molecules that are being made. On the other hand, if the random arrival of transcription factors at their target is dominant, we should have Eq (368) with the effective noise given by the Berg–Purcell limit,

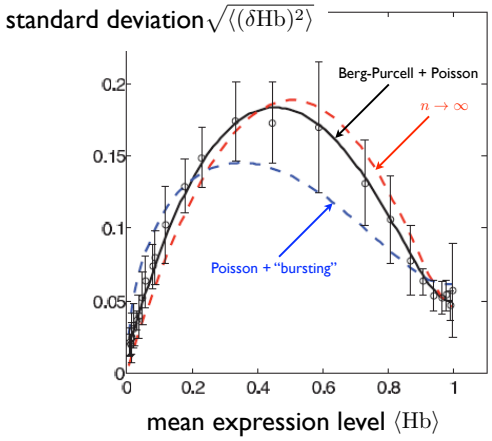

FIG. 55 Noise in Hunchback expression as a function of the mean expression level, from Tkačik et al (2008). This is a replotting of the data from Fig 54, compared with several models as described in the text. Error bars are standard deviations across multiple individual embryos.

so that

$$\langle(\delta\mathrm{Hb})^2\rangle_{\mathrm{BP}} = \left|\frac{d\langle\mathrm{Hb}\rangle}{d\ln c}\right|^2 \cdot \frac{1}{N_{\mathrm{cells}}Dac\tau_{\mathrm{avg}}}, \tag{370}$$

where we have added a factor to include, as above, the idea that Hb expression levels at one cell depend on an average over N_{cells} nearby cells. Empircally, the mean expression level is well approximated by a Hill function,

$$\langle\mathrm{Hb}\rangle = \frac{c^n}{c_{1/2}^n + c^n}, \tag{371}$$

where now we choose units where the maximum mean expression level is one, and the data are fit best by $n = 5$. Then we have

$$\frac{d\langle\mathrm{Hb}\rangle}{d\ln c} = n\langle\mathrm{Hb}\rangle\left(1 - \langle\mathrm{Hb}\rangle\right), \tag{372}$$

and hence, after some algebra,

$$\langle(\delta\mathrm{Hb})^2\rangle_{\mathrm{BP}} = \beta\langle\mathrm{Hb}\rangle^{2-1/n}\left(1 - \langle\mathrm{Hb}\rangle\right)^{2+1/n}, \tag{373}$$

$$\beta = \frac{n^2}{N_{\mathrm{cells}}Dac_{1/2}\tau_{\mathrm{avg}}}. \tag{374}$$

If we have both the Berg–Purcell noise at the input to transcriptional control, and the Poisson noise at the output, then we expect the variances to add, so that

$$\langle(\delta\mathrm{Hb})^2\rangle = \langle(\delta\mathrm{Hb})^2\rangle_{\mathrm{BP}} + \langle(\delta\mathrm{Hb})^2\rangle_{\mathrm{Poisson}} \tag{375}$$

$$= \beta\langle\mathrm{Hb}\rangle^{2-1/n}\left(1 - \langle\mathrm{Hb}\rangle\right)^{2+1/n} + \alpha\langle\mathrm{Hb}\rangle. \tag{376}$$

In Figure 55 we see how this prediction compares with experiment. Since $n = 5$ is known from the input/output relation, we have to set the parameters α and β. At maximal mean expression, $\langle\mathrm{Hb}\rangle = 1$ and Eq (376) predicts $\langle(\delta\mathrm{Hb})^2\rangle = \alpha$, so we can read this parameter directly from the behavior at the right hand edge of the graph ($\alpha^2 \sim 0.05$). We have just one parameter β left to fit, but this will determine the height, shape and position of the peak in the noise level vs mean, so it is not at all guaranteed that we will get a reasonable fit. In fact the fit is very good, and we find $\beta \sim 0.5$. It is interesting that the dependence of the variance on the mean seems very sensitive, since if we let the Hill coefficient become large, even the best fit of Eq (376) systematically misses the data, as shown by the $n \to \infty$ curve in Fig 55. Other subtly different models also fail, as you can see in Problem 65 [careful with number!].

Problem 64: Details of Hunchback noise. Discuss the meaning of the parameters α and β. Can you relate these to meaningful physical quantities? Do we have independent data to see if these numbers make sense?

Problem 65: Transcriptional bursting?. The key point about noise in synthesis and degradation is that we expect the variance to be monotonic as a function of the mean (as in the Poisson model), and this is not what we see in Fig 55. An alternative model that could explain the peak of noise at intermediate expression levels is that the transcription site switches between active and inactive states, generating a "burst" of mRNA molecules while in the active state. You should be able to go back to our discussion of noise in binding and unbinding without diffusion [leading up to Eq (A322)], and build up the predictions of this model.

(a.) Suppose that switching into the active state occurs at a rate $k_{\rm on}$, and the switch back to the inactive state occurs at a rate $k_{\rm off}$. These rates must vary with the concentration of the input transcription factor, since it is only by switching between active and inactive states that the system can modulate the mean output. It seems plausible that the mean output is proportional to the probability of being in the active state. Are there any conditions under which this would not be true?

(b.) Show that if the mean output is proportional to the probability of being in the active state, then the random switching will contribute to the output variance a term

$$\langle(\delta{\rm Hb})^2\rangle_{\rm burst} = \langle{\rm Hb}\rangle\,(1-\langle{\rm Hb}\rangle)\cdot\frac{\tau_c}{\tau_{\rm avg}}, \tag{377}$$

where the correlation time $\tau_c = 1/(k_{\rm on}+k_{\rm off})$, the output is measured in units such that the maximal mean value is $\langle Hb\rangle = 1$, as above, and we assume that the averaging time is long compared with τ_c.

(c.) Switching into the active state is associated with transcription factor binding. In contrast, switching back to the inactive state doesn't require any additional binding events. Thus it is plausible that the rate $k_{\rm off}$ is independent of the input concentration c. What is the dependence of $k_{\rm on}$ required to reproduce the mean input/output relation in Eq (371)? Is there a mechanistic interpretation of this dependence?

(d.) As an aside, can you give an alternative description based on the MWC model, as in our discussion of the bacterial rotary motor above? Notice that now you need to think about the kinetics of the transitions between the two states, not just the free energies. See also the Appendix A.4. This is deliberately open ended.

(e.) Combine your results in [b.] and [c.] to show that the analog of Eq (376) in this model is

$$\langle(\delta{\rm Hb})^2\rangle = \langle(\delta{\rm Hb})^2\rangle_{\rm burst} + \langle(\delta{\rm Hb})^2\rangle_{\rm Poisson} \tag{378}$$

$$= \gamma\langle{\rm Hb}\rangle\,(1-\langle{\rm Hb}\rangle)^2 + \alpha\langle{\rm Hb}\rangle. \tag{379}$$

Give an expression for γ in terms of the original parameters of the model. Explain why the steepness of the Hill function (that is, the parameter n) doesn't appear directly in determining the shape of the relation between variance and mean.

(f.) In Fig 55, we see the best fit of Eq (379) to the data, which is not very good. Without doing a fit, you should be able to show that the model predicts a relation between the point at which the noise is maximal, and the height of this maximum. Show that this is inconsistent with the data.

To summarize, we can now observe directly the noise in gene expression. While one could emphasize that these fluctuations are, under some conditions, quite large, it seems more surprising that there are conditions where they are quite small. Cells can set the output of their genetic control machinery with a precision of $\sim 10\%$ or better, thus doing much more than switching genes on and off—intermediate levels of expression are meaningful. This means, in particular, that we have make measurements with an accuracy of better than 10%, and this isn't always easy to do. More fundamentally, the precision with which cells can control expression levels is not far from the limits set by the random arrival of the relevant signaling molecules (transcription factors) at their targets. Of course, we could imagine cells which use more copies of all the transcription factors, and thus could achieve greater precision—or be sloppier, and reach the same precision—but this doesn't seem to be what happens. I don't think we understand why evolution has pushed cells into this particular corner.

So far we have discussed noise as a small fluctuation around the mean. It is also possible that, in the same way that thermal noise can result in a nonzero rate for chemical reactions, noise in chemical kinetics can generate spontaneous switching among otherwise stable states. Much has been written about this. I am less certain that we really understand any particular system. There is, however, some elegant physics here, so I would like to come back and discuss this.

The following two problems are concerned with a newly discovered bacterium that responds to a chemical signal by emitting light. The bacteria are roughly spherical, with diameter $d \sim 2\,\mu$m, and hence are clearly visible under the microscope. The chemical signal is shown to be a small protein, presumably secreted by other bacteria; the protein diffuses through the extracellular medium with a diffusion constant $D \sim 10\,\mu{\rm m}^2/{\rm s}$. Very careful experiments establish that each individual bacterium either emits light at full intensity or is essentially dark, and that changing the concentration c of the signaling protein changes the probability of being in the two states. Larger values of c correspond to higher probabilities of being in the light emitting state, so that $p_{\rm light}(c)$ is monotonically increasing.

Problem 66: Extreme sensitivity, but slowly. There is a specific concentration $c = c_{1/2}$ of the signaling protein such $p_{\rm light}(c_{1/2}) = p_{\rm dark}(c_{1/2}) = 0.5$. When poised at $c = c_{1/2}$ the system switches back and forth between the two states spontaneously at a rate of ~ 1/hour. Remarkably, a change in c by just 10% is sufficient to shift the probabilities from $p_{\rm light} = 0.5$ to $p_{\rm light} = 0.9$ or $p_{\rm light} = 0.1$ when the concentration is increased or decreased, respectively.

(a.) After some confusion in early experiments, it is found that everything said above is true, but the half–maximal concentration $c_{1/2} = 10^{-12}\,$M. Is this possible? Justify your answer clearly and quantitatively.

(b.) One group proposes that this extreme sensitivity is not at all surprising, since after all proteins can bind to other proteins with dissociation constants as small as $K_D \sim 10^{-15}\,$M. Does this observation of very tight binding have anything to do with the physical limits on sensitivity? Why or why not?

(c.) Another group notes that $10^{-12}\,$M corresponds to $\sim 10^{-3}$ molecules in the volume of the bacterium. They argue that this provides evidence for homeopathy, in which drugs are claimed to retain their effectiveness at extreme dilution, perhaps even to the point where the doses contain less than one molecule on average. Can you resolve their confusion?

Problem 67: How simple can it be? Further studies of this new light emitting bacterium aim at identifying the molecules involved. The first such experiment shows that if you block protein

synthesis, the system cannot switch between the dark and light states, indicating that the switch involves a change in gene expression rather than (for example) a change in phosphorylation or methylation states of existing proteins as in chemotaxis. A systematic search which knocks out individual genes, looking for effects on the behavior, finds only one gene that codes for a DNA–binding protein. When this gene is knocked out, all bacteria are permanently dark. More detailed experiments show that these bacteria not only are dark, they actually are not expressing the proteins required for generating light.

(a.) Draw the simplest schematic model suggested by these results. Be sure that your model explains why there are two relatively stable states (light and dark) rather than a continuum of intermediates, and that your model is consistent with the knock out experiments.

(b.) Assume that the signaling protein binds to some receptor on the surface of the cell and that this triggers a cascade of biochemical events. For simplicity you can imagine that the output of this cascade is some molecule, the concentration of which is proportional to the average occupancy of the receptors over some window of time. Explain how this molecule can couple to your model in [a] to influence the probability of the cell being in the dark or light states.

(c.) Formalize your models from [a] and [b] by writing differential equations for the concentrations of all the relevant species. Show how these equations imply the existence of discrete light/dark states. Can you see directly from the equations why changing the receptor occupancy will shift the balance between these states? It might be hard to explain the behavior near the midpoint ($c = c_{1/2}$), but it should be possible to explain the dominance of the dark state as $c \to 0$ and the light state as $c \to \infty$.

(d.) Describe qualitatively all the sources of noise that could enter your model. Do you have any guidance from experiment about which sources are dominant?

(e.) Consider the point where $c = c_{1/2}$. Explain qualitatively what features of your model are responsible for determining the ~ 1 hour time scale for jumping back and forth between the light and dark states.

(f.) See how far you can go in turning your remarks in [e] into an honest calculation!

There are several messages which I hoped to convey in this section. First, bacterial chemotaxis provides us with an example of chemical sensing which is interesting, not just in itself but as an example of a vastly more general phenomenon. Importantly, experiments on chemotaxis set a quantitative standard that should be emulated in the exploration of other chemical signaling systems, from the embryo to the brain. Second, as explained in Appendix A.6, the intuitive argument of Berg and Purcell can be made rigorous. What they identified is a limit to chemical signaling which is very much analogous to the photon shot noise limit in vision or imaging more generally. While molecules do many complicated things, they have to reach their targets in order to do them, and this is a random process, so this randomness sets a limit to the precision of almost everything that cells do.[47] Finally, real cells operate close to this limit, not just in specialized tasks such as chemotaxis but in the everyday business of regulating gene expression. While other noise sources are clearly present, the "noise floor" that results from the Berg–Purcell limit never seems far away, and in some cases cells may push all the way to the point where this is the dominant noise source.

The study of chemotaxis has a long history. From a biologist's point of view, the modern era starts when Adler (1965, 1969) demonstrates, using mutants, that chemosensing is independent of metabolism. From a physicist's point of view, the modern era starts when Berg builds his tracking microscope and observes, quantitatively, the paths of individual bacteria (Berg 1971, Berg & Brown 1972). The experiments which demonstrated the temporal character of the computations involved in chemotaxis were done by Macnab & Koshland (1972) and by Brown & Berg (1974). A nice discussion of how these temporal comparisons translate into mobility up the gradient of attractive chemical is given by Schnitzer et al (1990).

Adler 1965: Chemotaxis in *Escherichia coli*. *Cold Spring Harbor Symp Quant Biol* **30,** 289–292 (1965).

Adler 1969: Chemoreceptors in bacteria. J Adler, *Science* **166,** 1588–1597 (1969).

Berg 1971: How to track bacteria. HC Berg, *Rev Sci Instrum* **42,** 868–871 (1971).

Berg & Brown 1972: Chemotaxis in *Escherichia coli* analyzed by three–dimensional tracking. *Nature* **239,** 500–504 (1972).

Brown & Berg 1974: Temporal stimulation of chemotaxis in *Escherichia coli Proc Nat'l Acad Sci (USA)* **71,** 1388–1392 (1974).

Macnab & Koshland 1972: R Macnab & DE Koshland, The gradient–sensing mechanism in chemotaxis. *Proc Nat'l Acad Sci (USA)* **69,** 2509–2512 (1972).

Schnitzer et al 1990: Strategies for chemotaxis. M Schnitzer, SM Block, HC Berg & EM Purcell, *Symp Soc Gen Microbiol* **46,** 15–34 (1990).

For fluid mechanics in general, see Landau and Lifshitz (1987). The fact that bacteria live at low Reynolds number, and that this must matter for their lifetsyle, was surely was known to many people, for many years. But Berg's experiments on *E coli* provided a stimulus to think about this, and it resulted in a beautiful exposition by Purcell (1977), which has been hugely influential. The appreciation that self–propulsion at low Reynolds number has a gauge theory

[47] It is possible to produce light that does not obey Poisson statistics for the photon counts, and this raises the question of whether we could generate comparable noise reductions in chemical processes. I think the answer is yes—for example, one could transport molecules to their targets by an active process that is more orderly than diffusion—but this seems enormously costly, as first emphasized by Berg and Purcell themselves. It is, however, worth thinking about. More subtly, some chemical reactions involve enormous numbers of steps, so that the fractional variance in the time required for completion of the reaction by one molecule becomes very small, as in the discussion of rhodopsin shutoff in Section I.C. Indeed, transcription itself can be seen as an example, where it is possible for the time required to synthesize a single mRNA molecule—once transcription has been initiated—to be nearly deterministic, so that this process does not contribute a significant amount of noise.

description is due to Shapere & Wilczek (1987). The dramatic discovery that bacteria swim by rotating their flagella was made by Berg & Anderson (1973), and then Silverman & Simon (1974) succeeded in tethering cells by their flagella to see the rotation of the cell body.

Berg & Anderson 1973: Bacteria swim by rotating their flagellar filaments. *Nature* **245,** 380–382 (1973).

Landau & Lifshitz 1987: *Fluid Mechanics.* LD Landau & EM Lifshitz (Pergamon, Oxford, 1987).

Purcell 1977: EM Purcell, Life at low Reynolds' number, *Am J Phys* **45,** 3–11 (1977).

Shapere & Wilczek 1987: Self–propulsion at low Reynolds number. *Phys Rev Lett* **58,** 2051–2054 (1987).

Silverman & Simon 1974: Flagellar rotation and the mechanism of bacterial motility. M Silverman & M Simon, *Nature* **249,** 73–74 (1974).

Should add some references about rotation of the mitochondrial ATPase, and more recent work on flagellar motor ... .

:

The classic, intuitive account of the physical limits to chemcial sensing is by Berg and Purcell (1977). [Do we want to dig into the papers that they reference, in relation to sensitivity?] Measurements on the impulse response of the system were reported by Block et al (1982), and these experiments, along with Segall et al (1986) provide a more compelling demonstration that bacterium is sensitive to single molecular events. Another interesting paper from this period is Block et al (1983) [should tell the story about the Appendix as an example of models/theories in biology]. The idea of deriving the impulse response as the solution to an optimization problem, in the spirit of the Berg–Purcell discussion but more rigorously, has been explored by several groups: Strong et al (1998), Andrews et al (2006), and most recently Celani & Vergassola (2010), who introduced a novel game theoretic approach [check other refs].

Andrews et al 2006: Optimal noise filtering in the chemotactic response of *Escherichia coli.* BW Andrews, T–M Yi & PA Iglesias, *PLoS Comp Bio* **2,** e154 (2006).

Berg & Purcell 1977: Physics of chemoreception. HC Berg & EM Purcell, *Biophys J* **20,** 193–219 (1977).

Block et al 1982: Impulse responses in bacterial chemotaxis. SM Block, JE Segall & HC Berg, *Cell* **31,** 215–226 (1982).

Block et al 1983: Adaptation kinetics in bacterial chemotaxis. SM Block, JE Segall & HC Berg, *J Bateriol* **154,** 312–323 (1983).

Celani & Vergassola 2010: Bacterial strategies for chemotaxis. A Celani & M Vergassola, *Proc Nat'l Acad Sci (USA)* **107,** 1391–1396 (2010).

Segall et al 1986: Temporal comparisons in bacterial chemotaxis. JE Segall, SM Block & HC Berg, *Proc Nat'l Acad Sci (USA)* **83,** 8987–8991 (1986).

Strong et al 1998: Adaptation and optimal chemotactic strategy in *E coli.* SP Strong, B Freedman, W Bialek & R Koberle, *Phys Rev E* **57,** 5604–5617 (1998).

The experiments on the response of the flagellar motor to the CheY∼P concentration are by Cluzel et al (2000). For measurements on the diffusion constant of proteins in *E coli* see Elowitz et al (1999), and for observations on the structure of the motor in relation to its regulation by CheY∼P, see Thomas et al (1999). The model in Fig 49 is based on give original refs for MWC–style description of rotation. Give refs to models at the front end of the transduction scheme, depending on what gets said in the text!

Cluzel et al 2000: An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. P Cluzel, M Surette & S Leibler, *Science* **287,** 1652–1655 (2000).

Elowitz et al 1999: Protein mobility in the cytoplasm of *Escherichia coli.* MB Elowitz, MG Surette, P–E Wolf, JB Stock & S Leibler, *J Bateriol* **181,** 197–203 (1999).

Thomas et al 1999: Rotational symmetry of the C ring and a mechanism for the flagellar rotary motor. DR Thomas, DG Morgan & DJ DeRoiser, *Proc Nat'l Acad Sci (USA)* **96,** 10134–10139 (1999).

This seems to be the first place where GFP–based methods have come up, so need to give a guide ot the literature here!

:

In thinking about transcriptional regulation, it is useful to review some basic facts about molecular biology, for which the classic reference is Watson's *Molecular Biology of the Gene.* This has been through many editions, and at times flirted with being more of an encyclopedia than a textbook. I'll reference the current edition here, which seems a bit more compact than some of the intermediate editions, but I also encourage you to look back at earlier editions, written by Watson alone. A beautiful account of gene regulation, using the bacteriophage λ as an example, was given by Ptashne (1986), which has also evolved with time (Ptashne 1992); see also Ptashne (2001).

Ptashne 1986: *A Genetic Switch: Gene Control and Phage* λ. M Ptashne (Cell Press, Cambridge MA, 1986).

Ptashne 1992: *A Genetic Switch, Second Edition: Phage* λ *and Higher Organisms.* M Ptashne (Cell Press, Cambridge MA, 1992).

Ptashne 2001: *Genes and Signals.* M Ptashne (Cold Spring Harbor Laboratory Press, New York, 2001).

Watson et al 2008: *Molecular Biology of the Gene, Sixth Edition.* JD Watson, TA Baker, SP Bell, A Gann, M Levine &R Losick (Benjamin Cummings, 2008).

In order to make our discussion quantitative, we need to know the absolute concentration at which transcription factors act. Ptashne's books give some discussion of this, although the estimates were a bit indirect. Several groups have made measurements on the binding of transcription factors to DNA, trying to measure the concentration at which binding sites are half occupied; sometimes this is done by direct physical–chemical methods *in vitro*, and sometimes by less direct methods *in vivo*. Examples include Oehler et al (1994), Ma et al (1996), Pedone et al (1996), Burz et al (1998), and Winston et al (1999). A modern version of the *in vitro* binding experiment examines the molecules one at a time, as in the work by Wang et al (2009).

Burz et al 1998: Cooperative DNA binding by Bicoid provides a mechanism for threshold dependent gene activation in the *Drosophila* embryo. DS Burz, R Pivera–Pomar, H Jackle & SD Hanes, *EMBO J* **17,** 5998–6009 (1998).

Ma et al 1996: The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively. X Ma, D Yuan, K Diepold, T Scarborough, & J Ma, *Development* **122,** 1195–1206 (1996).

Oehler et al 1994: Quality and position of the three lac operators of *E coli* define efficiency of repression. S Oehler, M Amouyal, P Kolkhof, B von Wilcken–Bergmann & B Müller–Hill, *EMBO J* **13,** 3348–3355 (1994).

Pedone et al 1996: The single Cys2–His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high–affinity specific DNA binding. PV Pedone, R Ghirlando, GM Clore, AM Gronenborn, G Felsenfeld & JG Omichinski, *Proc Nat'l Acad Sci (USA)* **93,** 2822–2826 (1996).

Wang et al 2009: Quantitative transcription factor binding kinetics at the single molecule level. Y Wang, L Guo, I Golding, EC Cox, NP Ong, *Biophys J* **96,** 609–620 (2009).

Winston et al 1999: Characterization of the DNA binding properties of the bHLH domain of Deadpan to single and tandem sites. RL Winston, DP Millar, JM Gottesfeld, Gottesfeld & SB Kent. *Biochemistry* **38,** 5138–5146 (1999).

An important development in the field has been the construction of fusion proteins, combining transcription factors with fluorescent proteins, and the re–insertion of these fusions into the genome. For more about these techniques in general, see the references at the end of Section II.B. When cells divide, their contents are partitioned, and one can observe the noise from the finite number of molecules being assigned at random to one of the two daughter cells. Rosenfeld et al (2005), and more recently Teng et al (2010) has shown how this can be used to make very precise estimates of the number of copies of the protein in the mother cell, and thus providing a calibration that converts fluorescence intensity back into copy number. Gregor et al (2007a) discuss a case where it was possible to test in detail that the fusion construct replaces the function of the original transcription factor, quantitatively, and in the next paper they exploit this construct to analyze the noise in one step of transcriptional regulation (see below), as well as making estimates of absolute concentration by comparing the fluorescence intensity to a purified standard (Gregor et al 2007b).

Gregor et al 2007a: Stability and nuclear dynamics of the Bicoid morphogen gradient. T Gregor, EF Wieschaus, AP McGregor, W Bialek & DW Tank, *Cell* **130,** 141–152 (2007).

Gregor et al 2007b: Probing the limits to positional information. T Gregor, DW Tank, EF Wieschaus & W Bialek, *Cell* **130,** 153–164 (2007).

Rosenfeld et al 2005: Gene regulation at the single cell level. N Rosenfeld, JW Young, U Alon, PS Swain & MB Elowitz, *Science* **307,** 1962–1965 (2005).

Teng et al 2010: Measurement of the copy number of the master quorum–sensing regulator of a bacterial cell. S–W Teng, Y Wang, KC Tu, T Long, P Mehta, NS Wingreen, BL Bassler & NP Ong, *Biophys J* **98,** 2024–2031 (2010).

In contrast to bacteria, many eukaryotic cells are large enough, or move slowly enough, that they can get a reliable signal by measuring gradients across the length of their body; for a discussion of the limits to these measurements and some of the relevant experiments, see Endres & Wingreen (2009a,b). Need to digest data on chemotaxis in bigger cells ... . Find general reference on axon guidance, growth cones etc.. The measurements on extreme precision of axon guidance were reported by Rosoff et al (2004).

Endres & Wingreen 2009a: Accuracy of direct gradient sensing by single cells. RG Endres & NS Wingreen, *Proc Nat'l Aca Sci (USA)* **105,** 15749–15754 (2008).

Endres & Wingreen 2009b: Accuracy of direct gradient sensing by cell–surface receptors. RG Endres & NS Wingreen, *Prog Biophys Mol Biol* **100,** 33–39 (2009).

Gregor et al 2010: The onset of collective behavior in social amoebae. T Gregor, K Fujimoto, N Masaki & S Sawai, *Science* **328,** 1021–1025 (2010).

Rossof et al 2004: A new chemotaxis assay shows the extreme sensitivity of axons to molecular gradients. WJ Rosoff, JS Urbach, MA Esrick, RG McAllister, LJ Richards & GJ Goodhill, *Nature Neurosci* **7,** 678–682 (2004).

Song et al 2006: *Dictyostelium discoideum* chemotaxis: Threshold for directed motion. L Song, SM Nadkarnia, HU Bödeker, C Beta, A Bae, C Franck, W-J Rappel, WF Loomis & E Bodenschatz, *Eur J Cell Bio* **85,** 981–989 (2006).

It is only in the last decade that it has been possible to make direct measurements of the noise in gene expression, and even more recently that it has been possible to focus on noise in the control process itself. The initial experiment separating intrinsic from extrinsic noise sources using the two color plasmid was by Elowitz et al (2002), which touched off a series of experiments on both bacterial (Ozbudak et al 2002, Pedraza & van Oudenaarden 2005) and eukaryotic systems (Blake et al 2003, Raser & O'Shea 2004). The experiments on noise in the Bcd/Hb system are by Gregor et al (see above). A review of methods for measuring Bcd concentration profiles is given by Morrison et al (2011), and in particular they discuss the comparison of live GFP–based imaging with antibody staining methods in fixed samples. A more detailed analysis of the data on Bcd/Hb noise is given by Tkačik et al (2008), which also provides a broader context on the role of different noise sources in the control of gene expression. Models based on transcriptional bursting are inspired by the direct observation of these bursts in *E coli* by Golding et al (2005). It is worth thinking about whether the observed bursts necessarily result from the kinetics of switching between states of the transcriptional apparatus, or could be traced to the binding and unbinding of transcription factors.

Blake et al 2003: Noise in eukaryotic gene expression. WJ Blake, M Kaern, CR Cantor & JJ Collins, *Nature* **422,** 633–637 (2003).

Elowitz et al 2002: Stochastic gene expression in a single cell. MB Elowitz, AJ Levine, ED Siggia & PD Swain, *Science* **297,** 1183–1186 (2002).

Golding et al 2005: Real–time kinetics of gene activity in individual bacteria. I Golding, J Paulsson, SM Zawilski & EC Cox, *Cell* **123,** 1025–1036 (2005).

Morrison et al 2011: Quantifying the Bicoid morphogen gradient in living embryos. AH Morrisson, M Scheeler, J Dubuis & T Gregor, in *Imaging in Developmental Biology: A Laboratory Manual*, J Sharpe & R Wong, eds (Cold Spring Harbor Press, Woodbury NY, 2011); arXiv.org:1003.5572 [q–bio.QM] (2010).

Ozbudak et al 2002: Regulation of noise in the expression of a single gene. E Ozbudak, M Thattai, I Kurtser, AD Grossman & A van Oudenaarden, *Nature Gen* **31,** 69–73 (2002).

Pedraza & van Oudenaarden 2005: Noise propagation in gene networks. J Pedraza & A van Oudenaarden, *Science* **307,** 1965–1969 (2005).

Raser & O'Shea 2004: Control of stochasticity in eukaryotic gene expression. JM Raser & EK OShea, *Science* **304,** 1811–1814 (2004).

Tkačik et al 2008: The role of input noise in transcriptional regulation. G Tkačik, T Gregor & W Bialek, *PLoS One* **3,** e2774 (2008).

Will need to add some references about bistability, noise induced switching, and maybe path integral methods for noise .. depends on what gets said in the text.

C. More about noise in perception

We have already said a bit about noise in visual perception, in the case where perception amounts to counting photons. But this is just one corner of our perceptual experience, and we'd like to know if some of the same principles are relevant outside of this limit. In this section we will look at a few instances, sampled from different organisms and different sensory modalities. I think one of the important ideas here is that considerations of noise—and processing strategies for reaching reliable

conclusions in the presence of noise, perhaps even optimizing performance—cut across these many different systems, which often are the subjects of quite isolated literatures.

It has been known for some time that bats navigate by generating ultrasonic calls and listening for the echoes, forming an image of their world much as in modern sonar. To get a feeling for the precision of this behavior, there is a simple, qualitative experiment that is best explained with a certain amount of (literal) hand waving [ask Jim Simmons for original reference]. Some bats will happily eat mealworms if you toss them into the air. Before tossing them, however, you can dip them into a little bit of flour. To eat the worm, the bat must "see" it, and then maneuver its own body into position, finally sweeping the worm up in its wing and bringing it to its mouth. But if the worm has been dusted with flour, this will leave a mark on the wing. Now repeat the experiment, many times, with same bat (but, of course, different worms). If you look at the bat's wing, you might expect to see many spots of flour, but in fact all the spots are on top of one another. This suggests that the entire process—not just identifying the location of the worm in the air, but the acrobatic movements required to scoop it up—have a precision of roughly one centimeter. In echolocation, position estimates are based on the time delays of the echoes, and with a sound speed of $\sim 340\,\mathrm{m/s}$, this corresponds to a timing precision of $\delta t \sim 30\,\mu$s. This rough estimate already is interesting, although maybe not too shocking since we can detect a few microseconds of difference in the arrival times of sounds between our two ears, and this is how we can localize the source of low frequency sounds. Barn owls do even better, detecting $\delta\tau \sim 1\,\mu$sec between their ears.

As an aside, it was Rayleigh who understood that our brains need to use different cues for localization in different frequency ranges, just because of the physics of sound waves. At high frequencies (short wavelengths) our head casts an acoustic shadow, and there is a difference in intensity between out ears—the sound comes from the side that gets the louder signal. But at low frequencies, the wavelength is comparable to or larger than the size of our head, and there is no shadow. There is, however, a time or phase difference, but this is small. To demonstrate our sensitivity to these small time differences directly, he sat Lady Rayleigh in the gazebo behind their home, and arranged for tubes of slightly different length to lead from a sound source to her two ears. A fabulous image.

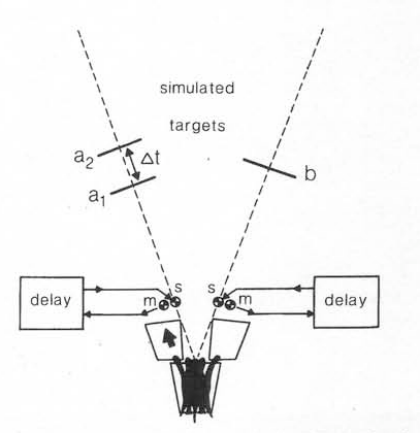

FIG. 56 A schematic of the 'Y' apparatus for testing echo timing discrimination performance in bats, from Simmons et al (1990).

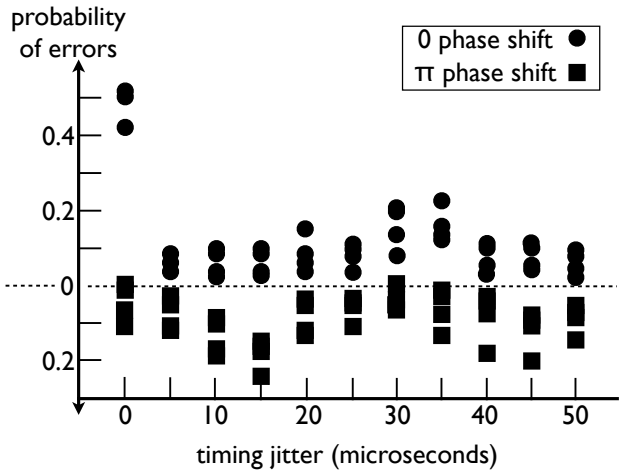

FIG. 57 Performance of four different bats at echo jitter discrimination, from Simmons et al (1990). Echoes can be returned with no phase shift (circles), or with a phase shift of π (squares); errors for the phase shifted echoes are measured downward.We see that the phase shift itself is detectable with almost no errors, that there is confusion around $\delta\tau \sim 35\,\mu$s, and that this "confusion peak" shifts and splits with the introduction of a phase shift.

Problem 68: Time differences and binaural hearing. Show that when a sound source is far away, the difference in propagation time to your two ears is independent of distance to the source. What does determine this time difference? For your own head, what is the time difference for a source at an angle of $\sim 10^\circ$ to the right of the direction your nose is pointing?

To be more quantitative, one would like to get the bats to report more directly on their estimates of echo delay, as in Fig 56. In one class of experiments, bats stand at the base of a Y with loudspeakers on the two arms. Their ultrasonic calls are monitored by microphones and returned through the loudspeakers with programmable delays. In

a typical experiment, the 'artificial echoes' produced by one side of the Y are at a fixed delay τ, while the other side alternately produces delays of $\tau \pm \delta\tau$. The bat is trained to take a step toward the side which alternates, and the question is how small we can make $\delta\tau$ and still have the bat make reliable decisions. Early experiments suggested that delays differences of $\delta\tau \sim 1\,\mu$sec were detectable, and perhaps more surprisingly that delays of $\sim 35\,\mu$sec were less detectable, as shown in Fig 57. The latter result might make sense if the bat were trying to measure delays by matching the detailed waveforms of the call and echo, since these sounds have most of their power at frequencies near $f \sim 1/(35\,\mu\text{sec})$—the bat can be confused by delay differences which correspond to an integer number of periods in the acoustic waveform, and one can even see the $n = 2$ 'confusion resonance' if one is careful. One can also introduce a phase shift into the artificial echo, and this shifts the confusion peak as expected.

Let's think about this more formally. Suppose that we are expecting a sound (or any signal) that has a time dependence $s_0(t)$, but we don't know when it will arrive, so what we actually observe will be $s_0(t-\tau)$ embedded in some background of noise. That is,

$$s(t) = s_0(t-\tau) + \eta(t), \tag{380}$$

where $\eta(t)$ is the noise. Let's assume, for simplicity, that the noise is white, with some spectral density $\mathcal{N}$. Then, as explained in Appendix B, the probability density for the function $s(t)$ becomes

$$P[s(t)|\tau] = \frac{1}{Z}\exp\left[-\frac{1}{2\mathcal{N}}\int dt\left|s(t)-s_0(t-\tau)\right|^2\right], \tag{381}$$

where Z is a normalization constant and the notation reminds us that this is the distribution if we know the delay τ. If instead the delay is $\tau+\delta\tau$,

$$P[s(t)|\tau+\delta\tau] = \frac{1}{Z}\exp\left[-\frac{1}{2\mathcal{N}}\int dt\left|s(t)-s_0(t-\tau-\delta\tau)\right|^2\right]. \tag{382}$$

As in our previous discussions of discrimination between two alternatives [give specific pointer], when we are faced with a particular signal $s(t)$ and have to decide whether the delay was τ or $\tau+\delta\tau$, the relevant quantity is the (log) likelihood ratio:

$$\lambda[s(t)] \equiv \ln\left(\frac{P[s(t)|\tau+\delta\tau]}{P[s(t)|\tau]}\right) \tag{383}$$

$$= -\frac{1}{2\mathcal{N}}\int dt\left|s(t)-s_0(t-\tau-\delta\tau)\right|^2 + \frac{1}{2\mathcal{N}}\int dt\left|s(t)-s_0(t-\tau)\right|^2 \tag{384}$$

$$= \frac{1}{\mathcal{N}}\int dt\, s(t)\left[s_0(t-\tau-\delta\tau)-s_0(t-\tau)\right]. \tag{385}$$

If the delay really is τ, then

$$\langle\lambda[s(t)]\rangle_\tau \equiv \left\langle\frac{1}{\mathcal{N}}\int dt\, s(t)\left[s_0(t-\tau-\delta\tau)-s_0(t-\tau)\right]\right\rangle_\tau \tag{386}$$

$$= \left\langle\frac{1}{\mathcal{N}}\int dt\,\left[s_0(t-\tau)+\eta(t)\right]\left[s_0(t-\tau-\delta\tau)-s_0(t-\tau)\right]\right\rangle_\tau \tag{387}$$

$$= \frac{1}{\mathcal{N}}\int dt\, s_0(t-\tau)\left[s_0(t-\tau-\delta\tau)-s_0(t-\tau)\right] \tag{388}$$

$$= \frac{1}{\mathcal{N}}[C(\delta\tau)-C(0)], \tag{389}$$

where

$$C(t) = \int dt'\, s_0(t')s_0(t'-t) \tag{390}$$

is the autocorrelation function of the expected signal. Similar calculations yield

$$\langle\lambda[s(t)]\rangle_{\tau+\delta\tau} = \frac{1}{\mathcal{N}}[C(0)-C(\delta\tau)], \tag{391}$$

$$\langle(\delta\lambda[s(t)])^2\rangle_\tau = \langle(\delta\lambda[s(t)])^2\rangle_{\tau+\delta\tau} \tag{392}$$

$$= \frac{2}{\mathcal{N}}[C(0)-C(\delta\tau)]. \tag{393}$$

It should also be clear that $\lambda[s(t)]$ is a Gaussian random variable (inherited from the Gaussian statistics of the noise η), so these few moments provide a complete description of the problem of discriminating between delays τ and $\tau + \delta\tau$. The end result is that the discrimination problem is exactly that of a single Gaussian variable (λ), with signal–to–noise ratio

$$SNR = \frac{(\langle\lambda[s(t)]\rangle_{\tau+\delta\tau} - \langle\lambda[s(t)]\rangle_\tau)^2}{\langle(\delta\lambda[s(t)])^2\rangle} = \frac{2}{\mathcal{N}}[C(0)-C(\delta\tau)]. \quad (394)$$

Thus we see that the SNR is large as soon as the jitter $\delta\tau$ is big enough to break the correlations in the waveform, and conversely that the SNR falls if shifting by $\delta\tau$ brings the waveform back into correlation with itself, as will happen for an approximately periodic signal such as the echolocation pulse.

Problem 69: Details of the SNR for detecting jitter in echolocation. Fill in the details leading to Eq (394).

(a.) How does this result change if the discrimination involves not just a time shift $\delta\tau$ but also a sign flip or π phase shift?

(b.) Recall the relationship between error probability and SNR [point back to photon counting discussion]. Is it practical to try and estimate the correlation function $C(\tau)$ by measuring the error probability as a function of $\delta\tau$? What if you also have access to experiments with a sign flip, as in (a.)? If you have errors in the measurement of the error probability, how do these propagate back to estimates of the underlying $C(\tau)$?

(c.) Compare your results in (b.) with the construction of "compound jitter discrimination curves" by Simmons et al (1990). Could you suggest improvements in their data analysis methods?

This argument about discriminability assumes that the bat's brain actually can compute using the entire acoustic waveform $s(t)$, rather some more limited features; in this sense we are describing the best that the bat could possibly do. It is interesting that such a calculation predicts confusion at delays where the autocorrelation function of the bat's call has a peak, and that such confusions are observed. On the other hand, this calculation seems hopelessly optimistic: "access to the acoustic waveform" means, in particular, access to features that are varying on the microsecond timescale. If we record the activity of single neurons emerging from the ear as they respond to pure tones, then we can see the action potentials "phase lock" to the tone, but this effect is significant only up to some maximum frequency. Beyond this high frequency cutoff, the overall rate of spikes increases with the intensity of the tone, but the timing of the spikes seems unrelated to the details of the acoustic waveform. Although there is controversy about the precise value of the cutoff frequency for phase locking, there seems to be no hint in the literature that it could be as high at 30 kHz. Taking all this at face value, it seems implausible that the auditory nerve actually transmits to the brain anything like a complete replica of the echo waveforms.

There is a second problem with this seemingly simple calculation. If we expand the SNR for small $\delta\tau$, we have

$$SNR = \frac{2}{\mathcal{N}}[C(0) - C(\delta\tau)] \approx \frac{C(0)}{\mathcal{N}} \cdot \left[\frac{C''(0)}{C(0)}\right] (\delta\tau)^2. \quad (395)$$

We expect that the term in brackets, which has the units of $1/(\text{time})^2$, is determined by the time scale on which the echolocation pulse is varying, something like $\sim 35\,\mu$sec. On the other hand, the first term, $C(0)/\mathcal{N}$ measures how loud the echo is relative to the background noise, and is dimensionless. We recall that in acoustics it is conventional to measure in deciBels, where 10 dB represents a factor of ten difference in acoustic power or energy. A typical quiet conversation produces sounds $\sim$ 30 dB above our threshold of hearing and hence above the limiting internal noise sources in the ear, whatever these may be. The bat's echolocation pulses are enormously loud, and although the echoes may be weak, it still is plausible that (at least in the laboratory setting) they are $\sim$ 60 dB above the background noise. This means that our calculation predicts a signal–to–noise ratio of one when the differences in delay $\delta\tau$ are measured in tens of *nanoseconds*, not microseconds. I think this was viewed as so obviously absurd that it was grounds for throwing out the whole idea that the bat uses detailed waveform information, even without reference to data on what the auditory nerve can encode.

In an absolutely stunning development, however, Simmons and colleagues went back to their experiments, produced delays in the appropriate range—convincing yourself that you have control of acoustic and electronic delays with nanosecond precision is not so simple—and found that the bats could do what they should be able to do as ideal detectors: they detect 10 *nano*second differences in echo delay, as shown in Fig 58. Further, they added noise in the background of the echoes and showed that performance of the bats tracked the ideal performance over a range of noise levels. This is a wonderful example with which to start this section of our discussion, since we have absolutely no idea how the bat manages this amazing feat of signal processing.

The problem of echo delay discrimination has just enough structure to emphasize an important point: when we make perceptual decisions, we are not identifying signals, we are identifying the distribution out of which these signals have been drawn. This becomes even more important as we move toward more complex tasks, where the randomness is intrinsic to the 'signal' rather than just a result of added noise. As an example, a single spoken word can generate a wide variety of sounds, all the more varied when embedded in a sentence. Identi-

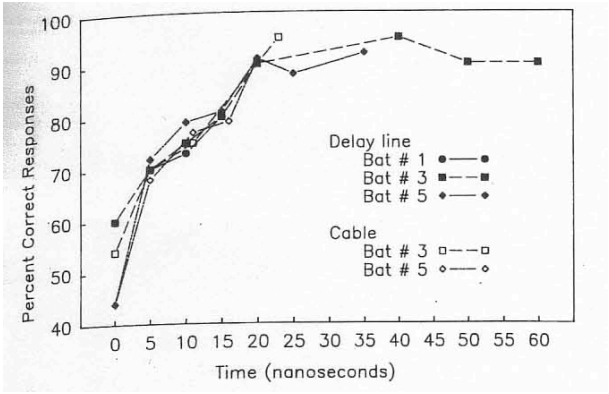

FIG. 58 Bat echo discrimination performance at very small delays, from Simmons et al (1990). Should add something about dependence on background noise level.]

fying the word really means saying that the particular sound we have heard comes from this distribution and not another. Importantly, probability distributions can overlap, and hence there are limits on the reliability of discrimination.

Some years ago, Barlow and colleagues launched an effort to use these ideas of discrimination among distributions to study progressively more complex aspects of visual perception, in some cases reaching into the psychology literature for examples of gestalt phenomena—where our perception is of the whole rather than its parts. One such example is the recognition of symmetry in otherwise random patterns. Suppose that we want to make a random texture pattern. One way to do this is to draw the contrast $C(\mathbf{x})$ at each point $\mathbf{x}$ in the image from some simple probability distribution that we can write down. An example is to make a Gaussian random texture, which corresponds to

$$P[C(\mathbf{x})] \propto \exp\left[-\frac{1}{2}\int d^2x \int d^2x' C(\mathbf{x})K(\mathbf{x}-\mathbf{x}')C(\mathbf{x}')\right], \tag{396}$$

where $K(\mathbf{x}-\mathbf{x}')$ is the kernel or propagator that describes the texture. By writing K as a function of the difference between coordinates we guarantee that the texture is homogeneous; if we want the texture to be isotropic we take $K(\mathbf{x}-\mathbf{x}') = K(|\mathbf{x}-\mathbf{x}'|)$. Using this scheme, how do we make a texture with symmetry, say with respect to reflection across an axis?

Problem 70: Texture discrimination. Show that Eq (396) can be rewritten as

$$P[C(\mathbf{x})] \propto \exp\left[-\frac{1}{2}\int \frac{d^2k}{(2\pi)^2}\frac{|\tilde{C}(\mathbf{k})|^2}{S_C(\mathbf{k})}\right], \tag{397}$$

where $S_C(\mathbf{k})$ is the (now two dimensional) power spectrum, connected as usual to the correlation function

$$\langle C(\mathbf{x})C(\mathbf{x}')\rangle = \int \frac{d^2k}{(2\pi)^2} S_C(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}. \tag{398}$$

Suppose that you have the task of discrimination between images drawn from distributions characterized by two different power spectra, $S_C(\mathbf{k})$ and $S_C(\mathbf{k}) + \Delta S_C(\mathbf{k})$. Show that, assuming one has access to a large area of the image, the discrimination problem for small $\Delta S_C(\mathbf{k})$ is again like the discrimination of a single Gaussian variable. Explain what role is played by the assumption of a "large area," and what defines large in this context. How does the signal–to–noise ratio for discrimination depend on area?

The statement that texture has symmetry across an an axis is that for each point $\mathbf{x}$ we can find the corresponding reflected point $\hat{\mathbf{R}}\cdot\mathbf{x}$, and that the contrasts at these two points are very similar; this should be true for every point. This can be accomplished by choosing

$$P_\gamma[C(\mathbf{x})] \propto \exp\left[-\frac{1}{2}\int d^2x \int d^2x' C(\mathbf{x})K(\mathbf{x}-\mathbf{x}')C(\mathbf{x}') + \frac{\gamma}{2}\int d^2x |C(\mathbf{x}) - C(\hat{\mathbf{R}}\cdot\mathbf{x})|^2\right], \tag{399}$$

where γ measures the strength of the tendency toward symmetry. Clearly as $\gamma \to \infty$ we have an exactly symmetric pattern, quenching half of the degrees of freedom in the original random texture. On the other hand, as $\gamma \to 0$, the weakly symmetric textures drawn from P_γ become almost indistinguishable from a pure random texture ($\gamma = 0$). Given images of a certain size, and a known kernel K, there is a limit to the smallest value of γ that can be distinguished reliably from zero, and we can compare this statistical limit to the performance of human observers. This is more or less what Barlow did, although he used blurred random dots rather than the Gaussian

textures considered here; the idea is the same, and all the details become the same in the limit of many dots. The result is that human observers come within a factor of two of the statistical limit for detecting γ or its analog in the random dot patterns. [Show a 1D version of this problem in a figure.]

One can use similar sorts of visual stimuli to think about motion, where rather than having to recognize a match between two halves of a possibly symmetric image we have to match successive frames of a movie. Here again human observers can approach the statistical limits, as long as we stay in the right regime: we seem not to make use of fine dot positioning (as would be generated if the kernel K only contained low order derivatives) nor can we integrate efficiently over many frames. These results are interesting because they show the potentialities and limitations of optimal visual computation, but also because the discrimination of motion in random movies is one of the places where people have tried to make close links between perception and neural activity in the (monkey) cortex.

Let us look in detail at the case of visual motion estimation, using not humans or monkeys, but a smaller system which have met once before—the visual system of the fly, which we have met already in Section I.A. If you watch a fly flying around in a room or outdoors, you will notice that flight paths tend to consist of rather straight segments interrupted by sharp turns and acrobatic interludes. These observations can be quantified through the measurement of trajectories during free flight, and in experiments where the fly is suspended from a torsion balance or a fine tether. Given the aerodynamics for an object of the fly's dimensions, even flying straight is tricky. In the torsion balance one can demonstrate directly that motion across the visual field drives the generation of torque, and the sign is such as to stabilize flight against rigid body rotation of the fly. Indeed one can close the feedback loop by measuring the torque which the fly produces and using this torque to (counter)rotate the visual stimulus, creating an imperfect 'flight simulator' for the fly in which the only cues to guide the flight are visual; under natural conditions the fly's mechanical sensors play a crucial role. Despite the imperfections of the flight simulator, the tethered fly will fixate small objects, thereby stabilizing the appearance of straight flight. Similarly, aspects of flight behavior under free flight conditions can be understood if flies generate torques in response to motion across the visual field, and that this response is remarkably fast, with a latency of just ~ 30 msec. The combination of free flight and torsion balance experiments strongly suggests that flies can estimate their angular velocity from visual input alone, and then produce motor outputs based on this estimate.

Voltage signals from the receptor cells are processed by several layers of the brain, each layer having cells organized on a lattice which parallels the lattice of lenses visible from the outside of the fly. As shown in Fig 59, after passing through the lamina, the medulla, and the lobula, signals arrive at the lobula plate. Here there is a stack of about 50 cells which are are sensitive to different components of motion. These cells have imaginative names, such as H1 and V1, which respond to horizontal and vertical components of motion, respectively. If one kills individual cells in the lobula plate then the simple experiment of moving a stimulus and recording the flight torque no longer works, strongly suggesting that these cells are an obligatory link in the pathway from the retina to the flight motor. Taken together, these observations support a picture in which the fly's brain uses photoreceptor signals to estimate angular velocity, and encodes this estimate in the activity of a few neurons.[48] What

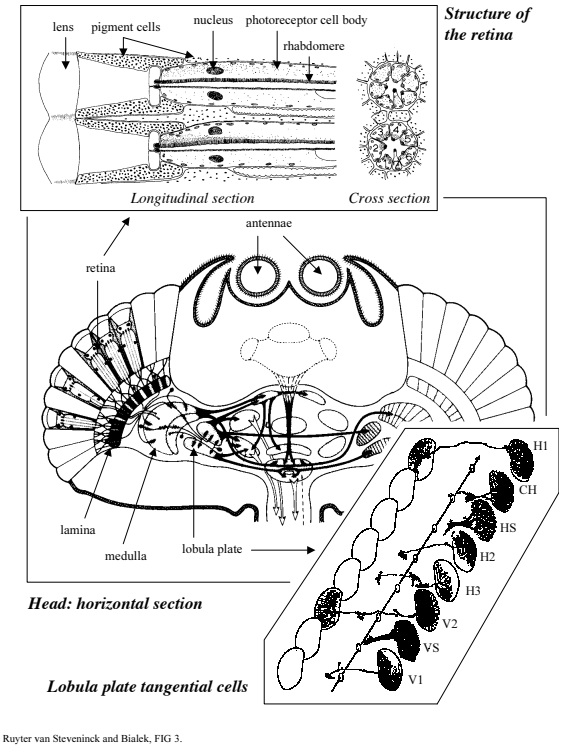

FIG. 59 The visual system of a fly, from the retina to the motion sensitive cells of the lobula plate. From de Ruyter van Steveninck & Bialek (2002).

[48] You should be skeptical of any claim about what the brain computes, or more generally what problems an organism has to solve in order to explain some observed behavior. The fact that flies can stabilize their flight using visual cues, for example, does *not* mean that they compute motion in any precise sense—they could use a form of 'bang–bang' control that needs knowledge only of the algebraic sign of the velocity, although I think that the tor-

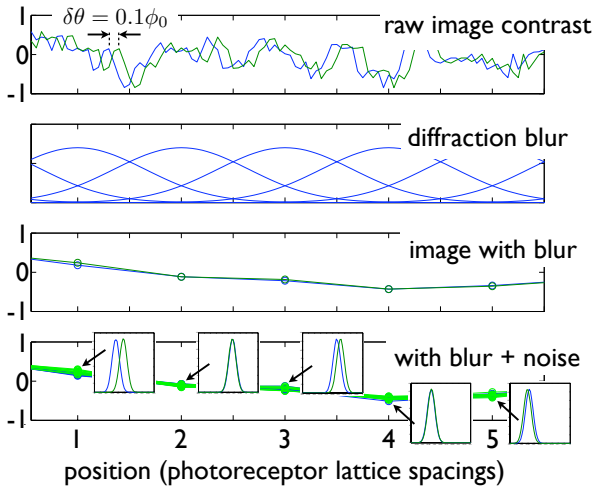

FIG. 60 The limits to motion detection. At top, a possible pattern of contrast (normalized light intensity) vs. position or angle in the visual world. Blue denotes the original pattern, and green illustrates a shift by one tenth of the spacing between photoreceptors. The second panel from the top shows the blurring and sampling of the image, with Gaussian apertures that provide a model for the optics of the fly's eye. Note that the spacing between photoreceptors is comparable to width of the diffraction blur. The third panel shows the signal arriving at each photoreceptor. We see that the blurring reduces the contrast enormously. The bottom panel illustrates the effect of adding noise, here with an amplitude expected if each snapshot involves counting an average of 10^3 photons. Insets show the distribution of signals plus noise in response to the original (blue) and shifted (green) images. Despite the large differences between the two initial patterns, only one of the five receptor cells shown here would be able to come near to reliable detection. The experiments described in the text are done under conditions of even smaller signal–to–noise ratios.

can we say about the physical limits to the precision of this computation?

Suppose that we look at a pattern of typical contrast C and it moves by an angle $\delta\theta$, as schematized in Fig 60. A single photodetector element will see a change in contrast of roughly $\delta C \sim C \cdot (\delta\theta/\phi_0)$, where ϕ_0 is the angular scale of blurring due to diffraction. If we can measure for a time τ, we will count an average number of photons $R\tau$, with R the counting rate per detector, and hence the noise can be expressed as a fractional precision in intensity of $\sim 1/\sqrt{R\tau}$. But fractional intensity is what we mean by contrast, so $1/\sqrt{R\tau}$ is really the contrast noise in one photodetector. To get the signal–to–noise ratio we should compare the signal and noise in each of the $N_{\rm cells}$ detectors, then add the squares if we assume (as for photon shot noise) that noise is independent in each detector while the signal is coherent:

$$SNR \sim N_{\rm cells} \cdot \left(\frac{\delta\theta}{\phi_0}\right)^2 C^2 R\tau. \qquad (400)$$

Motion discrimination is hard for flies because they have small lenses and hence blurry images (ϕ_0 is large) and because they have to respond quickly (τ is small); typical photon counting rates in a laboratory experiment are $R \sim 10^4\,{\rm s}^{-1}$ and outside on a bright day one can get to $R \sim 10^6\,{\rm s}^{-1}$. Under reasonable laboratory conditions—and taking account of all the factors that go in front of our rough Eq (400) in a more careful calculation—the optimal estimator would reach $SNR = 1$ at an angular displacement of $\delta\theta \sim 0.05°$.

We can test the precision of motion estimation in two very different ways. One is similar to the experiments we have discussed already, where we are forced to choose between two alternatives and measure the reliability of this choice. A single neuron responds to sudden steps of motion with a brief volley of action potentials which we

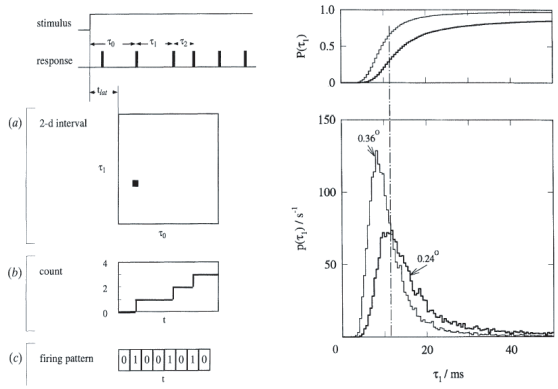

FIG. 61 Motion discrimination with the fly's H1 neuron, from de Ruyter van Steveninck & Bialek (1995). At left, a schematic of the spikes in response to a transient stimulus, such as a step of motion. We can describe the response by the time until the first spike τ_0, the time from the first spike to the second τ_1, ... Alternatively we can just count the spikes that have occurred up to a certain time after the stimulus, or we could at some fixed time resolution describe the whole pattern of spikes as a binary word. In each case we can analyze the discriminability of different stimuli by accumulating, over many repeated presentations of each stimulus, the distribution of responses. At right, an example of this analysis, focusing on the single interspike interval τ_1 in response to steps that differ in size by 0.12°. Long intervals correspond to the weaker stimulus, and from the cumulative probability distributions in the top panel we can read off the probabilities of correct identification of each stimulus.

sion balance experiments argue against such a model. It also is a bit mysterious why we find neurons with such understandable properties: one could imagine connecting photoreceptors to flight muscles via a network of neurons in which there is nothing that we could recognize as a motion–sensitive cell. Thus it is not obvious either that the fly must compute motion or that there must be motion–sensitive neurons.

can label as occurring at times $t_1, t_2, \cdots$. We as observers of the neuron can look at these times and try to decide whether the motion had amplitude θ_+ or θ_-; the idea is exactly the same as in earlier discussions of discrimination of signal vs noise, but here we have to measure the relevant probability distributions rather than making assumptions about their form; see Fig 61. Doing the integrals, one finds that looking at spikes generated in the first ~ 30 msec after the step (as in the fly's behavior) we can reach the reliability expected for $SNR = 1$ at a displacement $\delta\theta = |\theta_+ - \theta_-| \sim 0.12^\circ$, within a factor of two of the theoretical limit set by noise in the photodetectors.

It is worth noting a few more points that emerge from Fig 61 and further analyses of this experiment. First, on the ~ 30 msec time scale of relevance to behavior, there are only a handful of spikes. This is partly what makes it possible to do the analysis so completely, but it also is a lesson for how we think about the neural representation of information in general. Second, we can dissect the contributions of individual spikes to show that each successive spike makes a nearly independent contribution to the signal to noise ratio for discrimination, so there is essentially no redundancy. Finally, the motions we are discussing—motions close to the physical limits of detectability, and motions that real neurons can represent reliably—are much smaller than the lattice spacing on the retina or the nominal "diffraction limit" of angular resolution $\sim 1^\circ$. Analogous phenomena have been known in human vision for more than a century, and are called hyperacuity.

The step discrimination experiment gives us a very clear view of reliability in the neural response, but as with the other discrimination experiments discussed above it's not a very natural task. An alternative is to ask what happens when the motion signal (angular velocity $\dot{\theta}(t)$) is a complex function of time. Then we can think of the signal to noise ratio in Eq. (400) as being equivalent to a spectral density of displacement noise $N_\theta^{\rm eff} \sim \phi_0^2/(N_{\rm cells}C^2R)$, or a generalization in which the photon counting rate is replaced by an effective, frequency dependent, rate related to the noise characteristics of the photoreceptors, as in Fig 13. It seems likely, as discussed above, that the fly's visual system really does make a continuous or running estimate of the angular velocity, and that this estimate is encoded in the sequence of discrete spikes produced by neurons like H1. It is not clear that any piece of the brain ever "decodes" this signal in an explicit way, but if *we* could do such a decoding we could test directly whether the accuracy of our decoding reaches the limiting noise level set by noise in the photodetectors.

Decoding spike trains, at least under certain conditions, is much easier than one might have expected. The idea, shown in Fig 62, is that each spike contributes a small transient blip to our estimate of the signal vs. time, and to obtain the full estimate we add up all these small

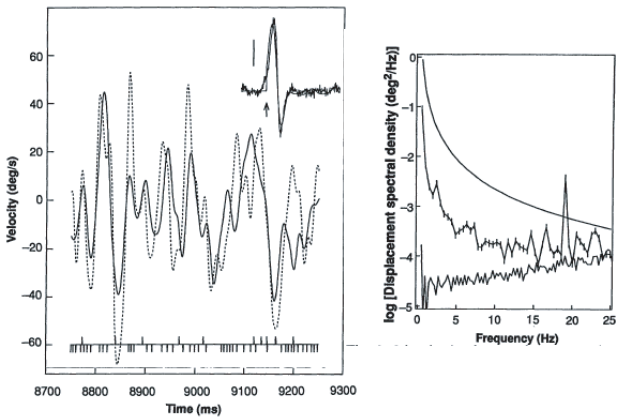

FIG. 62 Decoding continuous motion signals from spikes generated by the H1 neuron, from Bialek et al (1991). At left, dashed curve indicates the true stimulus, angular velocity as a function of time; solid line is the result of the decoding process, from Eq (403). Tick marks below the stimulus indicate the spikes generated in a single presentation of this stimulus (downward ticks) or its negative (upward ticks). This consideration of a hypothetical neuron that sees the negative stimulus is meant to restore symmetry between positive and negative velocities and corresponds roughly to the response of the H1 neuron on the other side of the fly's head, which has the opposite direction selectivity. At right is the spectral density of errors in the reconstruction. The error is reported as a displacement error, so the spectrum grows as $1/\omega^2$ for low frequencies. Also shown is the spectrum of the stimulus (smooth line) and the limiting noise level computed from the actual noise levels measured in fly photoreceptors under the same conditions as for these experiments on H1. Reconstruction error and the physical limit to precision converge at high frequencies, so that the fly approaches optimal performance.

contributions. Thus, if the signal we are interested in is $s(t)$, our estimate is

$$s_{\rm est}(t) = \sum_{\rm i} f(t - t_{\rm i}), \tag{401}$$

where $t_{\rm i}$ are the spike arrival times as before, and we can choose the filter $f(t)$ to minimize the errors

$$\chi^2 \equiv \int dt \left| s(t) - s_{\rm est}(t) \right|^2. \tag{402}$$

Like most neurons, H1 has a sign preference for its inputs—motion in one direction generate more spikes, while motion in the opposite direction generates fewer spikes. Thus, large negative velocities cause H1 to go silent, and in these periods we would have no basis for inferring the detailed waveform of velocity vs. time. Fortunately, the fly has two H1 neurons, one on each side of

the head, with opposite direction preferences. We could record from both cells, or we could use the fact that the two cells see opposite motions relative to their own preference, and look at the responses of one neuron to both a stimulus and the opposite motion. If the spikes in these two cases are $\{t_i^+\}$ and $\{t_i^-\}$, we can make a more symmetric reconstruction

$$s_{\rm est}(t) = \sum_i \left[f(t - t_i^+) - f(t - t_i^-) \right]. \tag{403}$$

Again, we choose the filter $f(t)$ to minimize χ^2.[49]

In Figure 62 we see that the reconstruction of the velocity waveform in fact is quite accurate. More quantitatively, the power spectrum of the errors in the reconstructed signal approaches the limit set by noise in the photoreceptor cells, within a factor of two at high frequencies. Further, one can change, for example, the image contrast and show that the resulting error spectrum scales as expected from the theoretical limit.

To the extent that the fly's brain can estimate motion with a precision close to the theoretical limit, we know that the act of processing itself does not add too much noise. But being quiet is not enough: to make maximally reliable estimates of nontrivial stimulus features like motion one must be sure to do the correct computation. Making this idea precise is in the same spirit as the discussion, in Section I.D, of pooling single photon signals from multiple rod cells at the level of bipolar cells. There we saw how the different orders of nonlinearity and summation result in very different final signal–to–noise ratios, even though all we are trying to do is add. Here the problem is more difficult, because the fly wants to estimate a feature of the visual world which is not directly reflected in the signals of any single receptor cell.

Problem 71: (Relatively) simple estimation problems. Suppose that someone draws a random number x from a probability distribution $P(x)$. Rather than seeing x itself, you get to see only a noisy version, $y = x + \eta$, where η is drawn from a Gaussian distribution with variance σ^2, so that

$$P(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y-x)^2 \right]. \tag{404}$$

Having seen y, your job is to estimate x.

(a.) Show that everything you know about x by virtue of observing y can be written in a way that suggests an analogy with statistical mechanics,

$$P(x|y) = \frac{1}{Z(y)} \exp\left[-\frac{V_{\rm eff}(x)}{k_B T_{\rm eff}} + \frac{F_{\rm eff} x}{k_B T_{\rm eff}} \right], \tag{405}$$

where

$$\frac{V_{\rm eff}(x)}{k_B T_{\rm eff}} = -\ln P(x) + \frac{x^2}{2\sigma^2} \tag{406}$$

$$k_B T_{\rm eff} = \sigma^2 \tag{407}$$

$$F_{\rm eff} = y. \tag{408}$$

(b.) From the discussion in Section I.D, we know that if we define "best" to be the estimator that minimizes χ^2, then the best estimator is the conditional mean,

$$x_{\rm est}(y) = \int dx\, x P(x|y). \tag{409}$$

Construct $x_{\rm est}(y)$ in the case where $P(x)$ is a Gaussian with unit variance. Show that this estimate, although "best," is systematically wrong. That is, if we average $x_{\rm est}(y)$ over the distribution $P(y|x)$, we do not recover x itself. Explain why this can still be the best estimate.

(c.) Now consider the case $P(x) = (1/2)\exp(-|x|)$. Show that, even though the transformation from what we are interested in (x) to what we measure (y) is linear, the optimal estimator is nonlinear. In particular, if rather than asking for an estimator that minimizes χ^2, we ask for the most probable value of x given y, show that the optimal estimator involves a threshold nonlinearity.

Motion estimation is an example of the more general problem of perceptual estimation. The data to which the brain has access are the responses of receptor cells, and the goal is to estimate some feature of the world. The first key step is to use Bayes' rule, combining the noisy data from the receptors with our prior knowledge that some things are more likely than others. Schematically,

$$P(\text{feature}|\text{receptor responses}) = \frac{P(\text{receptor responses}|\text{feature})P(\text{feature})}{P(\text{receptor responses})}. \tag{410}$$

The second key step is to note that receptors typically don't respond directly to the features of interest, but rather to raw sensory signals such as light intensity, sound pressure in the auditory system, the concentrations of specific molecular species in complex odors, etc.. Continuing schematically, let's denote the full spatiotemporal pattern of light intensities falling on the retina by $\mathcal{I}$. Receptor responses really depend on $\mathcal{I}$, which in turn is correlated with the feature that we want to estimate. Thus,

$$P(\text{receptor responses}|\text{feature}) = \int D\mathcal{I}\, P(\text{receptor responses}|\mathcal{I}) P(\mathcal{I}|\text{feature}), \tag{411}$$

and putting all the terms together we have

$$P(\text{feature}|\text{receptor responses}) = \frac{1}{P(\text{receptor responses})} \int D\mathcal{I}\, P(\text{receptor responses}|\mathcal{I}) P(\mathcal{I}, \text{feature}). \tag{412}$$

If the lights are bright, and the noise level in the photoreceptors is low, it is plausible that knowing the pattern of receptor responses is almost equivalent to knowing the spatiotemporal pattern of light intensities $\mathcal{I}$, and hence viewed as a function of $\mathcal{I}$ the distribution $P(\text{receptor responses}|\mathcal{I})$ is very sharply peaked. Then the entire structure of the optimal computation that maps receptor responses to the desired feature is controlled by $P(\mathcal{I}, \text{feature})$, which is a property of the world that we live in rather than of our eyes or brains. This is perhaps our most important qualitative conclusion: optimal estimates of sensory features involve computations determined by the structure of the world around us. To the extent that our brains, and those of other animals, make optimal estimates, this means that the way in which we process the world is set by the physics of our environment, not by peculiarities of our biological hardware.

For the case of motion estimation, what is the structure of $P(\mathcal{I}, \text{feature})$? For simplicity let's think about a one–dimensional version of the problem, so that spatiotemporal pattern of light intensity $\mathcal{I} \equiv I(x,t)$. Then if a small piece of the visual world is moving rigidly relative to us with a velocity v, we should have $I(x,t) = I_0(x - vt)$. Then we can take derivates in space and time,

$$\frac{\partial I(x,t)}{\partial x} = I_0'(x - vt) \tag{413}$$

$$\frac{\partial I(x,t)}{\partial t} = -vI_0'(x - vt). \tag{414}$$

Thus, we can compute the velocity as a ratio of spatial and temporal derivatives,

$$v_{\rm est} = -\frac{\partial I(x,t)/\partial t}{\partial I(x,t)/\partial x}. \tag{415}$$

This is correct, but we have derived it by pushing to extremes. First we said that noise in the receptor responses is negligible, so we can say that we are effectively computing functions of the light intensity itself. Then we assumed that the dynamics of the light intensity is determined only by motion at the single velocity v. If either of these assumptions breaks down, our "gradient based" estimator of velocity, Eq (415) gets into serious trouble.

When we deal with noisy data we develop several intuitions. First, the nature of our measurements is such that there usually is relatively more noise at higher frequencies, both in time and in space. Thus, to suppress noise, we average. Conversely, if we differentiate, we expect that noise will be amplified, since differentiation enhances higher frequencies. Second, when we have a noisy measurement, it is dangerous to put this in the denominator of a ratio—there is a chance that we will divide by zero, because of a fluctuation. The gradient based estimator compounds these sins, differentiating and then taking a ratio. We expect that this will be a disaster if our low noise assumptions are violated.

Problem 72: Ratios of noisy numbers. Suppose that we have two numbers that we try to measure, a and b. Our measurements, which we can call $\hat{a}$ and $\hat{b}$, give us the values of a and b but with some added Gaussian noise, so that

$$P(\hat{a}|a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\hat{a}-a)^2/2\sigma^2}; \tag{416}$$

for simplicity we'll assume that the noise level is the same for our measurements of b, so that

$$P(\hat{b}|b) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\hat{b}-b)^2/2\sigma^2}. \tag{417}$$

What we would like to do is to estimate the ratio $r \equiv a/b$ from our measurements $\hat{a}$ and $\hat{b}$.

(a.) Suppose we make form a naive estimate just by taking the ratio of our measurements, $r_{\rm est}^{\rm naive} = \hat{a}/\hat{b}$. Do a small simulation to examine numerically the probability distribution of this estimate. In particular, consider the case where $a = b = 1$, so the correct answer is $r = 1$. If $\sigma = 0.1$, presumably $r_{\rm est}^{\rm naive}$ stays close to this correct answer, but what happens at $\sigma = 0.2$ or 0.5? How does the variance of the estimator $r_{\rm est}^{\rm naive}$ change as the noise level σ increases? Be sure to check in your simulation that you have enough samples to get a reliable measure of the variance. Is there anything suspicious in this computation, especially at larger σ?

(b.) Look more closely at the right hand tail of the distribution of $r_{\rm est}^{\rm naive}$, that is the behavior of $P(r_{\rm est}^{\rm naive} \gg 1)$ in the case where $a = b = 1$. Plot your numerical results on linear, semilog, and log–log plots to see if you can recognize the shape of the tail. If the shape changing with the noise level σ? Try to make a precise statement based on your simulations. I have left this somewhat open ended.

(c.) Try to derive analytically the regularities that you found in [b].

(d.) Although we think of $\hat{a}$ and $\hat{b}$ as measurements of the separate variables a and b, really all we want to know is the ratio $r \equiv a/b$. Show that the best estimate can be written, using Bayes' rule, as

$$r_{\rm est}(\hat{a},\hat{b}) = \int dr \frac{r}{P(\hat{a},\hat{b})} \int da \int db\, \delta\left(r - \frac{a}{b}\right) P(\hat{a}|a) P(\hat{b}|b) P(a,b). \tag{418}$$

Make as much progress as you can evaluating these integrals on the hypothesis that the prior distribution $P(a,b)$ is broad and featureless. If you want to proceed analytically, you may find it useful to introduce a Fourier representation of the delta function, and look for a saddle point approximation. Numerically, you could assume, for example, that $P(a,b)$ is uniform over some region of the $a - b$ plane, and just do the integrals for representative values of $\hat{a}$ and $\hat{b}$, mapping the function $r_{\rm est}(\hat{a},\hat{b})$. Can you verify that $r_{\rm est}^{\rm naive}$ is close to optimal at very small values of σ? What happens at larger values of σ? If σ is fixed, what happens as $b \to 0$?

The most obvious problem with the gradient motion estimator in Eq (415) is simply that it is not well defined when the spatial derivative becomes small. This problem exists even if noise in the photoreceptors is small. To address the problem we have to understand what the distribution $P(\mathcal{I}, \text{feature})$ looks like. Conceptually, what we want to do is simple. Imagine taking a walk on a very still day, so that motions of the world relative to our retina (or relative to the fly's retina) are dominated by our own motion. If we carry a camera as we walk, we can take a movie, and we can also put a gyroscope on the camera to monitor it's motion. What emerges from such an experiment, then, is a set of samples drawn out of the distribution $P(\mathcal{I}, \text{feature})$. In particular, pixel by pixel and moment by moment, we can compute the spatial and temporal derivatives in the movie, and measure the velocity as well, so that we sample the distribution $P(\partial I/\partial t, \partial I/\partial x, v)$.

If the gradient based estimate of motion were exact, then the distribution $P(\partial I/\partial t, \partial I/\partial x, v)$ would be very sharply peaks along a ridge where $v = -(\partial I/\partial t)/(\partial I/\partial x)$. To see if this is right, we can compute directly the optimal estimator. We know that the best estimate in the sense of χ^2 is the conditional mean, so should compute[50]

$$v_{\rm est}(\partial_t I, \partial_x I) = \int dv\, v \frac{P(\partial_t I, \partial_x I, v)}{P(v)}. \tag{419}$$

The results of this computation, based on a walk in the woods, are shown in Fig 63.[51] We see that, when the spatial gradients are large, the contours of constant $v_{\rm est}$ really are straight lines, as expected from the gradient based estimator. But when the spatial gradients are smaller, a new structure emerges, which is more closely approximated by a *product* of derivatives, $v_{\rm est} \propto (\partial I/\partial t) \times (\partial I/\partial x)$, rather than a ratio. As you can see in the following problem, the same product structure emerges if we go back to the general formulation and take the limit of high noise levels.

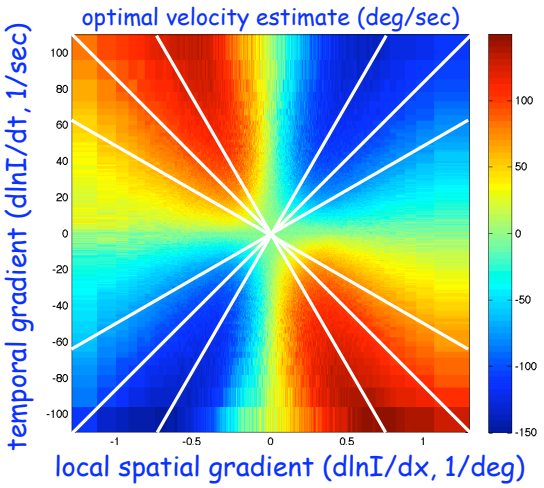

FIG. 63 Optimal estimates of angular velocity as a function of local spatial and temporal gradients of light intensity. Computed from the theory described in the text, with the joint distribution of movies and motion sampled experimentally. Images are collected through an optical system that matches the fly's eye, and smoothed in time with a filter that optimizes estimation performance. At small signals, near the center of the plot, we see that moving along a line of constant physical velocity (in white; $\partial_t I + v\partial_x I = 0$) results in a changing estimate—a systematic error; only for large signals is the optimal estimate veridical. Experiments by SR Sinha & RR de Ruyter van Steveninck.

Problem 73: Series expansion of the optimal estimator at low signal–to–noise ratios. We know from Section I.A that photoreceptors in the fly respond linearly to changes in light intensity or contrast [point back to specific equations; check consistency of notation]. If the fly is rotating relative to the world along an angular trajectory $\theta(t)$, then the spatiotemporal pattern of contrast (again in a one–dimensional model) is $C(x - \theta(t), t)$. Individual cells respond with voltages $V_{\rm n}(t)$ given by

$$V_{\rm n}(t) = \int dt'\, T(t - t') \int dx\, M(x - x_{\rm n}) C(x - \theta(t'), t'), \tag{420}$$

where $T(\tau)$ is the temporal impulse response function and $M(x - x_{\rm n})$ is an aperture function centered on a lattice point $x_{\rm n}$ in the retina.

(a.) Show that the distribution of all the voltages given the trajectory can be written as

$$P[\{V_{\rm n}(t)\}|\theta(t)] \propto \int DC\, P[C] \exp\left[-\frac{1}{2}\sum_{\rm n}\int \frac{d\omega}{2\pi} \frac{|\tilde{V}_{\rm n}(\omega) - \langle \tilde{V}_{\rm n}(\omega)\rangle|^2}{N_V(\omega)}\right], \tag{421}$$

where the mean voltages are, in the Fourier representation,

$$\langle \tilde{V}_{\rm n}(\omega)\rangle = \tilde{T}(\omega) \int dx\, M(x - x_{\rm n}) \int dt\, e^{+i\omega t} C(x - \theta(t), t), \tag{422}$$

$N_V(\omega)$ is the power spectrum of the voltage noise, and $P[C]$ is the distribution of contrast that the fly would observed if held at $\theta = 0$.

(b.) The optimal estimator is the conditional mean,

$$\dot{\theta}_{\rm est}(t_0) = \int D\theta\, \dot{\theta}(t_0) P[\theta(t)|\{V_{\rm n}(t)\}] \tag{423}$$

$$P[\theta(t)|\{V_{\rm n}(t)\}] = \frac{P[\{V_{\rm n}(t)\}|\theta(t)]P[\theta(t)]}{P[\{V_{\rm n}(t)\}]}. \tag{424}$$

Evaluate all the integrals in a perturbation series, assuming that the average voltage responses are small compared with the noise

[50] I need to make a segue between the notation $\partial I/\partial x$ and $\partial_x I$.

[51] Although conceptually simple, to generate Fig 63 requires measuring light intensities with spatial and temporal resolution matched to that of the retina, but collecting much more light so that photon shot noise in these measurements will be less than that in the retina and one can meaningfully claim to measure intensity at the input to the visual system. For details, as always, see the references at the end of the section.

level. You should find that the leading term is

$$\dot{\theta}_{\rm est}(t) \approx \sum_{\rm nm} \int d\tau \int d\tau' V_{\rm n}(t-\tau) K_{\rm nm}(\tau,\tau') V_{\rm m}(t-\tau'). \quad (425)$$

Relate the kernel $K_{\rm nm}(\tau,\tau')$ to expectation values in the distributions $P[C(x,t)]$ and $P[\theta(t)]$.

(c.) Can you reformulate the expansion so that instead of expanding for small overall signal–to–noise ratio (small R), you expand for small instantaneous signals, that is for small $V_{\rm n}(t)$? What happens to the kernels in this case? It seems obvious that there shouldn't be a linear term in this expansion. Can there be a third order term? If such a term exists, what happens to the optimal estimate of velocity when if we show the same movie, but with inverted contrast (exchanging black for white)?

We can understand the low signal to noise ratio limit by realizing that when something moves there are correlations between what we see at the two space–time points (x,t) and $(x+v\tau, t+\tau)$. These correlations extend to very high orders, but as the background noise level increases the higher order correlations are corrupted first, until finally the only reliable thing left is the two–point function, and closer examination shows that near neighbor correlations are the most significant: we can be sure something is moving because signals in neighboring photodetectors are correlated with a slight delay. This form of "correlation based" motion computation, schematized in Fig 64, was suggested long ago by Reichardt and Hassenstein based on behavioral experiments with beetles.

There are two clear signatures of the correlation model. First, since the receptor voltage is linear in response to image contrast, the correlation model confounds contrast with velocity: all things being equal, doubling the image contrast causes our estimate of the velocity to increase by a factor of four (!). This is an observed property of the flight torque that flies generate in response to visual motion, at least at low contrasts, and the same quadratic behavior can be seen in the rate at which motion sensitive neurons generate spikes, as shown in Fig 65. Even humans experience the illusion of contrast dependent motion perception at very low contrast. Although this might seem strange, it's been known for decades.

The second signature of correlation computation is that we can produce movies which have the right spatiotemporal correlations to generate a nonzero estimate $\dot{\theta}_{\rm est}$ but don't really have anything in them that we would describe as "moving" objects or features. Consider a spatiotemporal white noise movie $\psi(\mathbf{x},t)$,

$$\langle \psi(\mathbf{x},t)\psi(\mathbf{x}',t')\rangle = \delta(\mathbf{x}-\mathbf{x}')\delta(t-t'), \quad (426)$$

and then add the movie to itself with a weight and an offset:

$$C(\mathbf{x},t) = \psi(\mathbf{x},t) + a\psi(\mathbf{x}+\mathbf{\Delta x}, t+\Delta t). \quad (427)$$

Composed of pure noise, there is nothing really moving here. If you watch the movie, however, there is no question that you think it's moving, and the fly's neurons respond too (just like yours, presumably). Even more impressive is that if you change the *sign* of the weight a, then the direction of motion reverses, as predicted from the correlation model.

Problem 74: Motion from correlations alone. Generate the image sequences described in the previous paragraph, and verify that you (and your friends) perceive them as moving.

(a.) Play with the amplitude and sign of the weight a to see how it influences your perception. Can you find a regime in which the speed of motion seems to depend on $|a|$? Can you verify the reversal of motion when $a \to -a$?

(b.) Compute the correlation function $\langle C(\mathbf{x},t)C(\mathbf{x}',t')\rangle$; for simplicity you might want to confine your attention to a one dimensional example. Consider also the correlation function for a genuine moving image, in which $C(\mathbf{x},t) = C_0(\mathbf{x}-\mathbf{v}t)$. If $\mathbf{v} = \mathbf{\Delta x}/\Delta t$, how do the two correlation functions compare?

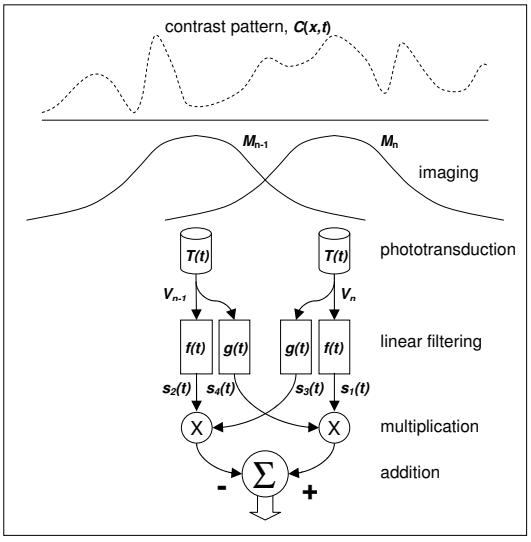

FIG. 64 The correlator model of visual motion detection, adapted from Reichardt (1961). A spatiotemporal contrast pattern $C(x,t)$ is blurred by the photoreceptor point spread function, $M(x)$, and sampled by an array of photoreceptors, two of which (neighboring photoreceptors numbers $n-1$ and n) are shown here. After phototransduction, the signals in each photoreceptor are filtered by two different linear filters, $f(t)$ and $g(t)$. The outputs of these filters from the different photoreceptors, $s_1(t)$ and $s_3(t)$ from photoreceptor n and $s_2(t)$ and $s_4(t)$ from photoreceptor $n-1$ are multiplied and one of these products is subtracted from the other by the addition unit, yielding a direction selective response. Thanks to Rob de Ruyter for this figure.

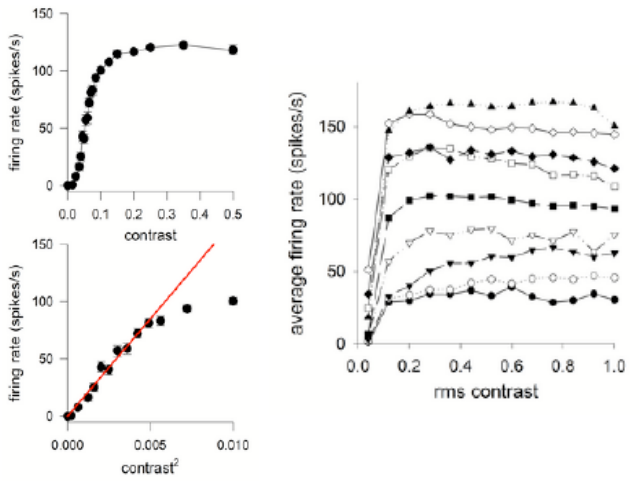

FIG. 65 Responses of the H1 neuron to moving scenes with varying contrast. Scenes consist of bars with random intensities, moving at constant velocity. At left, at one particular velocity we measure the rate at which H1 generates action potentials, as a function of contrast. Lower panel expands the region at low contrast, emphasizing the quadratic behavior. At right, the responses at multiple velocities, showing that the "saturated" response at high contrast still is sensitive to the speed of movement. [does this appear in a paper? details of stimuli?] Thanks to Rob de Ruyter for this figure.

The optimal motion estimator illustrates the general tradeoff between systematic and random errors. If we really are viewing an image that moves rigidly, so that $C(x,t) = C(x+vt)$, then there is no question that the "right answer" is to compute v as the ratio of temporal and spatial derivatives. Any departure from this involves making a systematic error. But, as discussed above, taking derivatives and ratios are both operations which are perilous in the presence of noise. To insulate the estimate from random errors driven by such noise (or, more generally, by aspects of the image dynamics that are not related to motion), we must calculate something which, typically, will not give the "right answer" even on average—we accept some systematic errors in order to reduce the impact of random errors. In the context of perception, systematic errors have a special name: illusions.

Could the theory of optimal estimation be a quantitative theory of illusions, grounded in physical principles? Colloquially, we say that "to err is human," and it is conventional to assume that cases in which biological systems get the wrong answer to their signal processing problems provide evidence regarding the inadequacies of the biological hardware. Is it possible that, rather than being uniquely human or biological, to err is the optimal response to the limits imposed by the physical world?

The long history of the correlation model provides ample testimony that insect visual systems make the kind of systematic errors expected from the optimal estimator, but precisely because of this long history it is hard to view these are successful predictions. It would be more compelling if we could show that the same system which is well described by the correlator under some conditions crosses over to something more like the ratio of derivatives model at high signal–to–noise ratio, but this has been elusive. The contrast dependence of the response in the motion sensitive neurons saturates at high contrast, and this saturated response still varies with velocity (Fig 65), as if the larger signals allow the system to disentangle ambiguities and recover a veridical estimate, but other experiments suggest that errors inherent in the correlation model persist even with strong signals. Humans easily see the illusion of motion with the noise movies of Eq (427), as well as other motion illusions, but at high signal–to–noise ratios our visual systems recover estimates of velocity which are not systematically distorted, suggesting that in primates there is some sort of crossover between different limits of the motion computation, and there are efforts to make the correspondence with the optimal estimator more quantitative. Experiments under more natural, free flight conditions show that both flies and bees have access to veridical estimates of their translational velocity and can use this to control their flight speed, in contrast to what one would have expected from the correlator model, and it worth noting that the responses of the motion–sensitive neurons are also very different under more natural conditions.

[This needs to be clearer] In Figure 66 we see the responses of the H1 neuron to the rotation of a fly, outside under nearly natural conditions. During the course

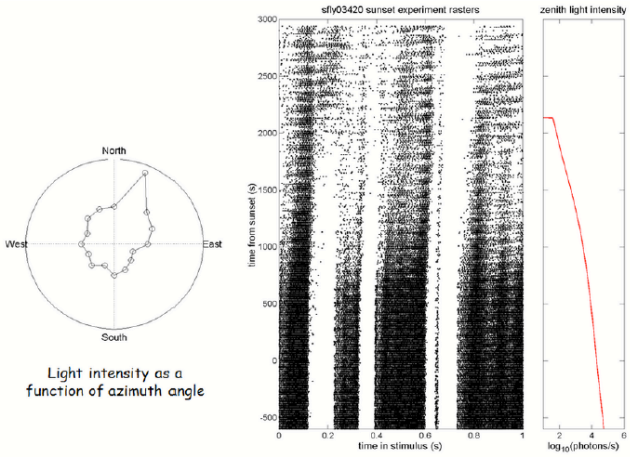

FIG. 66 Responses of the H1 neuron s a fly is rotated outside, over a period during which the mean light level is falling. [fill in the caption. does this appear in a paper? details of stimuli?] Thanks to Rob de Ruyter for this figure.

of the experiment, the sun was going down, and so the mean light level varied by several orders of magnitude as the same trajectory of angular velocity vs. time was repeated over and over. The integral of the trajectory was not quite zero, however, so that on each repetition the spatial pattern of light intensity was a bit different even if the angular velocity was the same. At the start of the experiment, the responses are extremely vigorous, and insensitive to the variations in the spatial structure of the visual environment. As the light level falls, responses become weaker, but more dramatically we see that there is a systematic variation from repetition to repetition, which appears as a diagonal pattern of spikes across the upper part of Fig 66. Thus, when signal–to–noise ratios are high in the natural environment, H1 responds to time dependent velocities and largely ignores the spatial structure of its environment, while at lower signal–to–noise ratios the confounding of spatial structure and motion becomes more and more obvious. This pattern is in agreement with the expectations from optimal estimation theory, according to which such systematic errors arise only from the need to insulate the computation from random noise.

What we would really like is to have methods of dissecting the computation that has been done by a neuron, simply by analyzing the relationship between visual inputs and spiking outputs under natural conditions. This is a huge challenge, and obviously would be interesting in many other contexts. Approaches to this problem are discussed in Appendix A.7, where we also see results that come closest to a smoking gun for the crossover between correlator and gradient computations.

For visual signal processing, getting our hands on the true distribution of signals in the natural environment is a difficult experiment. For seemingly more complex "cognitive" judgments, the situation, perhaps surprisingly, is much simpler. To give an example, suppose that you are told of a member of the United States Congress who has served for $t = 15$ years. What is your prediction for how long his total term will last? To keep things as simple as possible, let's assume you are not told anything about the politics of this congressman or his district; all you have to work with is $t = 15$ and your general knowledge of the turnover of elected officials. Obviously your knowledge is probabilistic, so we use Bayes' rule to write

$$P(t_{\rm total}|t) \propto P(t|t_{\rm total})P(t_{\rm total}). \qquad (428)$$

If the moment at which the question is asked is not somehow synchronized to the length of congressional terms, then we have to assume that $P(t|t_{\rm total})$ is uniform, $P(t|t_{\rm total}) = 1/t_{\rm total}$. Thus our inference is controlled by the "prior" distribution $P(t_{\rm total})$, and we can look this up in a database about the history of the congress. Finally, if you must pick one value of $t_{\rm total}$, it makes sense in this context to choose the median, the point at which the actual value of $t_{\rm total}$ is equally likely to be longer or shorter than your estimate. As an example, if $P(t_{\rm total})$ is a reasonably narrow Gaussian distribution, then for t much less than the mean $\langle t_{\rm total}\rangle$, our best estimate of $t_{\rm total}$ is just $\langle t_{\rm total}\rangle$ itself, while if the time t is much larger than the mean then our best estimate is only slightly higher than t, which makes sense. Other priors, of course, can give qualitatively different results.

Problem 75: Estimating $t_{\rm total}$. Derive the results just stated for the Gaussian prior. Consider also cases where $P(t_{\rm total}) \propto t_{\rm total}^{-\gamma}$ or $P(t_{\rm total}) \propto t_{\rm total}^{n} e^{-t_{\rm total}/\tau}$.

The example of congressional terms is not unique. We could ask, as insurance companies do (albeit with more input data), about human lifespans: if you meet someone of age t, what is your best guess about their life expectancy? If you make a phone call and have been on hold for t minutes, what is your best guess about the total time you will have to wait? If you find yourself on line t of a poem, what is your best guess about the total length of text? Nor is the structure of the problem bound to time, as such: suppose you learn that a movie has collected t dollars in gross receipts; what is your bets guess about what its total earnings will be? All these problems have in common that we can look up the correct distribution $P(t_{\rm total})$. Another important feature is that we can just go ask people what they think, and see how they do relative to the predictions for optimal estimation based

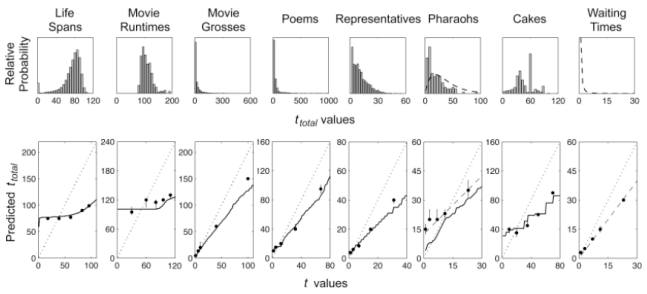

FIG. 67 Estimation of totals based on one observation, from Griffiths & Tenenbaum (2006). The top row shows the priors $P(t_{\rm total})$ measured from real world data. The bottom panel compares people's predictions (points) based on one observation t with the optimal median estimator (solid lines) and a naive "universal" estimate $\hat{t}_{\rm total} = 2t$. For the reigns of Pharaohs and the telephone waiting times, dashed lines show optimal estimators for $P(t_{\rm total}) \propto t_{\rm total} e^{-t_{\rm total}/\tau}$ ($\tau = 17.9$) and $P(t_{\rm total}) \propto t_{\rm total}^{-\gamma}$ ($\gamma = 2.43$), respectively.

on the priors appropriate to our real world. The results from such an experiment are shown in Fig 67.

I found the results of Fig 67 quite astonishing when I first saw them. The time it takes to bake a cake comes from a very irregular distribution, but people seem to know this distribution and estimate accordingly. They are a bit confused about how long the Pharoahs reigned, but their confusion is consistent: estimation of $t_{\rm total}$ behaves as if the subjects know the shape of $P(t_{\rm total})$ but are off on the mean time scale, and if you ask another group of subjects to guess the mean reign of the Pharoahs, they deviate from the right answer by the same factor. Important as the telephone problem may be, this is one case where there is no convenient data to which we can refer, so this case remains untested. In all the other cases, however, spanning seemingly very different domains of knowledge and very different shapes for $P(t_{\rm total})$, people are performing close to the optimum.

If we trace through the details of optimal estimation theory, one can see that construction of the correct estimator involves knowing not only the distribution of signals, but also the distribution of noise. Perhaps the simplest illustration of this is given by the problem of combining two measurements. Suppose that we are interested in x, but we observe

$$y_1 = x + \eta_1 \tag{429}$$
$$y_2 = x + \eta_2, \tag{430}$$

where the noise levels on the two measurements are generally different, $\langle \eta_1^2 \rangle = \sigma_1^2$ and $\langle \eta_2^2 \rangle = \sigma_2^2$; for simplicity we will assume that the noise is Gaussian. Intuitively, we should be able to do better by combining the two observations than we would do by looking just at one of them, and we also expect that we should give greater weight to the more accurate measurement. Quantitatively, if the measurements are independent of one another, we have

$$P(x|y_1,y_2) = \frac{P(y_1,y_2|x)P(x)}{P(y_1,y_2)} \tag{431}$$
$$\propto P(x)P(y_1|x)P(y_2|x) \tag{432}$$
$$\propto P(x)\exp\left[-\frac{1}{2\sigma_1^2}(y_1-x)^2 - \frac{1}{2\sigma_2^2}(y_2-x)^2\right]. \tag{433}$$

Then we can form the optimal estimator in the least squares sense,

$$x_{\rm est}(y_1,y_2) \equiv \int dx\, x P(x|y_1,y_2) \tag{434}$$
$$= \frac{\sigma_2^2 y_1 + \sigma_1^2 y_2}{\sigma_1^2+\sigma_2^2}, \tag{435}$$

where in the last step we assume that the prior $P(x)$ is broad compared with the noise levels in our data. Thus, as expected, the optimal estimate is a combination of the data, and the weights are inverse to their relative noise levels.

Problem 76: Cue combination. Fill in the details leading to Eq (435). Can you work out the same problem but with additional multiplicative noise, $y_{\rm n} = e^{g_{\rm n}} x + \eta_{\rm n}$, where $g_{\rm n}$ is also Gaussian? In this case, it is possible to generate errors that are very large, so presumably large disagreements between the data points y_1 and y_2 should not be resolved by simple averaging. See how much analytic progress you can make here, or do a simple simulation. This is deliberately open ended.

There are many situations in which we give strongly unequal weights to different data. A dramatic example is ventriloquism, in which we trust our eyes not our ears, and assign the source of speech to the person (or the dummy) whose lips are visibly moving. To see whether we are giving weights in relation to noise levels, as would be optimal, we have to do an experiment in which we can manipulate the effective noise levels. This was first done convincingly in tasks that require subjects to combine information from vision and touch, [add figure from Ernst & Banks, with explanation]. Although under normal conditions we give strong preference to our visual system, these data show convincingly that we do this only because our visual system provides much more accurate spatial information; if we can change their noise levels, people will change the weights given to different cues, as predicted by optimal estimation theory.

[loss functions, actions .. Maloney; Wolpert]

The examples of estimation that we have discussed thus far have in common that the distribution of the feature we are interested in estimating has a single well defined peak given the input sensory data. In many cases, however, the data that we collect with our senses have multiple interpretations, perhaps even multiple interpretations that provide equally good explanations of what we have seen or heard. These 'ambiguous percepts' arise in many contexts. When we experience these stimuli, our perceptions jump at random among the different possibilities. Could these random jumps originate from the same small noise sources that limit the reliability of our senses? [give fuller discussion, both visual and auditory examples ... alternative models ... maybe end with connection to conscious perception?]

Need to give a summary/conclusion for the section.

While there were many precursors, reaching back across centuries, the conclusive demonstration that bats navigate by echolocation, with sounds beyond the range of human hearing, was by Griffin & Galambos (1941). Griffin (1958) gives a beautiful presentation of the history and basic facts about the system. [need original ref for exp't with dusted mealworms] The first suggestion of sub–microsecond precision in this system was from Simmons (1979).

Perhaps not surprisingly, these observations (and the provocative title of the paper in which they were presented) touched off a flurry of controversy; for different views, see Altes (1981) and Menne & Hackbarth (1986). The astonishing results on nanosecond precision, and the optimality of performance in background noise, were presented by Simmons et al (1990). For context, it is interesting to look at examples of precise timing measurements in binaural hearing [need ref, presumably to Konishi in barn owls] and in weakly electric fish (Rose & Heiligenberg 1985).

Altes 1981: Echo phase perception in bat sonar? RA Altes, *J Acoust Soc Am* **69,** 1232–1246 (1981).

Griffin 1958: *Listening in the Dark.* DR Griffin (Yale University Press, New Haven, 1958).

Griffin & Galambos 1941: The sensory basis of obstacle avoidance by flying bats. DR Griffin & R Galambos, *J Exp Zool* **86,** 481–506 (1941).

Menne & Hackbarth 1986: Accuracy of distance measurement in the bat *Eptesicus fuscus*: Theoretical aspects and computer simulations. D Menne & H Hackbarth, *J Acoust Soc Am* **79,** 386–397 (1986).

Rose & Heiligenberg 1985: Temporal hyperacuity in the electric sense of fish. G Rose & W Heiligenberg, *Nature* **318,** 178–180 (1985).

Simmons 1979: Perception of echo phase information in bat sonar. JA Simmons, *Science* **204,** 1336–1338 (1979).

Simmons et al 1990: Discrimination of jittered sonar echoes by the echolocating bat, *Eptesicus fuscus*: The shape of target images in echolocation. JA Simmons, M Ferragamo, CF Moss, SB Stevenson & RA Altes, *J Comp Physiol A* **167,** 589–616 (1990).

The program of comparing human performance with statistical limits in the context of higher level perception was outlined by Barlow (1980). The experiments on symmetry in random dot patterns are by Barlow & Reeves (1979), and an analysis of optimality in motion perception using random dot stimuli was given by Barlow & Tripathy (1997). For a review of how these stimuli have been used to probe the connections between neural activity and perception, see Newsome et al (1995). [Probably there needs to be a bit more here (!); maybe also in the text?] Note that, as discussed in *Spikes* (Rieke et al 1997; see below), these experiments connecting neural activity with perception in primates have been done, largely, in a regime where the subject is integrating imperfectly over very long periods of time, much longer than we would expect to see constant velocity motion in a natural setting; see also Osborne at al (2004). This complicates efforts to compare either neural or behavioral performance with the physical limits, and indeed I don't know of any effort to measure the responses of visual cortex in a regime (e.g., photon counting in the dark) where we understand fully the sources of noise limiting our perception; there is an opportunity here.

Barlow 1980: The absolute efficiency of perceptual decisions. HB Barlow, *Phil Trans R Soc Ser B* **290,** 71–82 (1980).

Barlow & Reeves 1979: The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. HB Barlow & BC Reeves, *Vision Res* **19,** 783–793 (1979).

Barlow & Tripathy 1997: Correspondence noise and signal pooling in the detection of coherent visual motion. H Barlow & SP Tripathy, *J Neurosci* **17,** 7954–7966 (1997).

Newsome et al 1995: Visual motion: Linking neuronal activity to psychophysical performance. WT Newsome, MN Shadlen, E Zohary, KH Britten & JA Movshon, in *The Cognitive Neurosciences,* M Gazzaniga, ed, pp 401–414 (MIT Press, Cambridge, 1995).

Osborne et al 2004: Time course of information about motion direction in visual area MT of macaque monkeys. LC Osborne, W Bialek & SG Lisberger, *J Neurosci* **24,** 3210–3222 (2004).

The classical work on motion estimation in insect vision was by Hassenstein and Reichardt (1956); perspectives on these early ideas are given by Reichardt (1961) and by Reichardt and Poggio (1976). A crucial piece of data in this discussion concerns the speed of a flying insect's motor response to visual motion, and a first estimate of this was given by Land and Collett (1974) in a beautiful analysis of natural flight trajectories; subsequent work was done by Wagner (1986a–c) and by Schilstra and van Hateren (1999; van Hateren & Schilstra 1999).

van Hateren & Schilstra 1999: Blowfly flight and optic flow. II. Head movements during flight. JH van Hateren & C Schilstra, *J Exp Biol* **202,** 1491–1500 (1999).

Hassenstein & Reichardt 1956: Systemstheoretische Analyse der Zeit–, Reihenfolgen–, und Vorzeichenauswertung bei der Bwegungsperzeption des Rüsselkäfers. S Hassentsein & W Reichardt, *Z Naturforsch* **11b,** 513–524 (1956).

Land & Collett 1974: Chasing behavior of houseflies (*Fannia canicularis*): A description and analysis. MF Land & TS Collett, *J Comp Physiol* **89,** 331–357 (1974).

Reichardt 1961: Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. W Reichardt, in *Sensory Communication*, WA Rosenblith, ed, pp 303–317 (MIT Press, Cambridge 1961).

Reichardt & Poggio 1976: Visual control of orientation behavior in flies. I. A quantitative analysis. W Reichardt & T Poggio, *Q Rev Biophys* **9,** 311–375 (1976).

Schilstra & van Hateren 1999: Blowfly flight and optic flow. I: Thorax kinematics and flight dynamics. C Schilstra & JH van Hateren, *J Exp Biol* **202,** 1481–1490 (1999).

Wagner 1986a: Flight performance and visual control of flight in the free–flying house fly (*Musca domestica L.*). I: Organization of the flight motor. H Wagner, *Phil Trans R Soc Lond Ser B* **312,** 527–551 (1986).

Wagner 1986b: Flight performance and visual control of flight in the free–flying house fly (*Musca domestica L.*). II: Pursuit of targets. H Wagner, *Phil Trans R Soc Lond Ser B* **312,** 553–579 (1986).

Wagner 1986c: Flight performance and visual control of flight in the free–flying house fly (*Musca domestica L.*). I: Interactions between angular movement induced by wide– and small–field stimuli. H Wagner, *Phil Trans R Soc Lond Ser B* **312,** 581–595 (1986).

[Introduce this with refs to the anatomy of the fly visual system.] Motion sensitive neurons in the fly visual system were discovered by Bishop & Keehn (1966), around the same time that Barlow et al (1964) discovered motion sensitive neurons in the rabbit retina. Today we take for granted that individual neurons can be selective for very complicated things, culminating in face– and object–selective neurons in the far reaches of the visual cortex [refs to Gross et al], but these early measurements were surprising. Indeed, in Barlow's hands, the observation of motion sensitivity played a key role in helping to shape the idea that cells respond to successively more complex conjunctions of features as we move through successive layers of processing [need to find which of HBB's refs is best here]. An early experiment showing that some of the motion sensitive neurons are a necessary link in optomotor behavior is by Hausen & Wehrhahn (1983); [since then ... ?].

Barlow et al 1964: Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. HB Barlow, RM Hill & WR Levick, *J Physiol (Lond)* **173,** 377–407 (1964).

Bishop & Keehn 1966: Two types of neurones sensitive to motion in the optic lobe of the fly. LG Bishop & D G Keehn, *Nature* **212**, 1374–1376 (1966).

Hausen & Wehrhahn 1983: Microsurgical lesion of horizontal cells changes optomotor yaw responses in the blowfly *Calliphora erythrocephala*. K Hausen & C Wehrhahn, *Proc R Soc Lond B* **219**, 211–216 (1983).

The experiments on the precision of motion discrimination using the output of H1 are from de Ruyter van Steveninck & Bialek (1995), and the reconstruction of velocity waveforms was done in Bialek et al (1991); a review of these ideas and results is given in *Spikes* (Rieke et al 1997). A detailed calculation of the physical limits to motion estimation in this system is in my lecture notes from the Santa Fe Summer School (Bialek 1990). For a general discussion of hyperacuity in vision see Westheimer (1981), and for the relation of hyperacuity to physical limits, see Geisler (1984). The theory of optimal motion estimation is from Marc Potters' PhD thesis (Potters & Bialek 1994); related work was done by [need to understand exactly what Simoncelli and others did around the same time], and application of these ideas to human visual motion perception can be found in Weiss et al (2002). Problem [**] about third order statistics is inspired by Fitzgerald et al (2011). [Stocker?]

Bialek 1990: Theoretical physics meets experimental neurobiology. W Bialek, in *1989 Lectures in Complex Systems, SFI Studies in the Sciences of Complexity, Vol II,* E Jen, ed, pp 513–595 (Addison–Wesley, Menlo Park CA, 1990).

Bialek et al 1991: Reading a neural code. W Bialek, F Rieke, RR de Ruyter van Steveninck & D Warland, *Science* **252,** 1854–1857 (1991).

Fitzgerald et al 2011: Symmetries in stimulus statistics shape the form of visual motion estimators. JE Fitzgerald, AY Katsov, TR Clandinin & MJ Schnitzer, *Proc Nat'l Acad Sci (USA)* in press (2011).

Geisler 1984: Physical limits of acuity and hyperacuity. WS Geisler, *J Opt Soc Am A* **1,** 775–782 (1994).

Potters & Bialek 1994: Statistical mechanics and visual signal processing. M Potters & W Bialek, *J Phys I France* **4**, 1755–1775 (1994).

Rieke et al 1997: *Spikes: Exploring the Neural Code*. F Rieke, D Warland, R de Ruyter van Steveninck & W Bialek (MIT Press, Cambridge, 1997).

de Ruyter van Steveninck & Bialek 1995: Reliability and statistical efficiency of a blowfly movement–sensitive neuron. R de Ruyter van Steveninck & W Bialek, *Phil Trans R. Soc Lond Ser B* **348,** 321–340 (1995).

Weiss et al 2002: Motion illusions as optimal percepts. Y Weiss, EP Simoncelli & EH Adelson, *Nature Neurosci* **5,** 598–604 (2002).

Westheimer 1981: Visual hyperacuity. G Westheimer, *Prog Sens Physiol* **1,** 1–30 (1981).

The classical evidence for the systematic errors of motion estimation predicted by the correlator model are discussed by Reichardt & Poggio (1976), above. Experiments showing the quadratic contrast dependence of responses in the motion sensitive neurons include [need to find the early ones!]. The demonstration that quadratic behavior at low contrasts coexists with unambiguous responses to velocity at high contrast is given by de Ruyter van Steveninck et al (1994, 1996) [check that these are the best references!]. These experiments were done with randomly textured images, whereas classical studies of visual motion have used periodic gratings. The correlator model also predicts that velocity will be confounded with the spatial frequency of these gratings, and this error persists even under high signal–to–noise ratio conditions (Haag et al 2004); it is not clear whether this represents a genuine failure of optimal estimation, a byproduct of strategies for gain control and efficient coding (Borst 2007), or simply a behavior that would never be seen under natural conditions. There are several experiments, especially in bees (Srinivasan et al 1991, 1996; Baird et al 2005), indicating that insects have access to signals that allow them to control their flight speed without any of the systematic errors predicted by the correlator model; recent work confirms this conclusion in *Drosophila* using sophisticated tracking and virtual reality to allow control experiments under free flight conditions (Fry et al 2009). A number of experiments have shown that the responses of motion–sensitive neurons are also very different under more natural conditions (Lewen et al 2001, de Ruyter van Steveninck et al 2001), although most of the analysis has focused on the nature of coding in spike trains rather than the nature of the motion computation itelf. [Is Rob's experiment on reducing ambiguity at high light levels, outside, published?] An attempt to dissect the motion computation represented by the spiking output of H1 is described in Bialek & de Ruyter van Steveninck (2005), and in Appendix A.7. [Do we say something about controversies?]

Baird et al 2005: Visual control of flight speed in honeybees. E Baird, MV Srinivasan, S Zhang & A Cowling, *J Exp Biol* **208,** 3895–3905 (2005).

Bialek & de Ruyter van Steveninck 2005: Features and dimensions: Motion estimation in fly vision. W Bialek & R de Ruyter van Steveninck, arXiv:q–bio/0505003 (2005).

Borst 2007: Correlation versus gradient type motion detectors: the pros and cons. A Borst, *Phil Trans R Soc Lond Ser B* **362,** 369–374 (2005).

Fry et al 2009: Visual control of flight speed in *Drosophila melanogaster*. SN Fry, N Rohrseitz, AD Straw & MH Dickinson, *J Exp Biol* **212,** 1120–1130 (2009).

Lewen et al 2001: Neural coding of naturalistic motion stimuli. GD Lewen, W Bialek & RR de Ruyter van Steveninck, *Network* **12,** 317–329 (2001); arXiv:physics/0103088 (2001).

de Ruyter van Steveninck et al 1994: Statistical adaptation and optimal estimation in movement computation by the blowfly visual system. RR de Ruyter van Steveninck, W Bialek, M Potters & RH Carlson, in *Proc IEEE Conf Sys Man Cybern*, 302–307 (1994).

de Ruyter van Steveninck et al 1996: Adaptive movement computation by the blowfy visual system. RR de Ruyter van Steveninck, W Bialek, M Potters, RH Carlson & GD Lewen in *Natural and Artificial Parallel Computation: Proceedings of the Fifth NEC Research Symposium*, DL Waltz, ed, 21–41 (SIAM, Philadelphia, 1996).

de Ruyter van Steveninck et al 2001: Real time encoding of motion: Answerable questions and questionable answers from the fly's visual system. R de Ruyter van Steveninck, A Borst & W Bialek, in *Processing Visual Motion in the Real World: A Survey of Computational, Neural and Ecological Constraints,* JM Zanker & J Zeil, eds, pp 279–306 (Springer–Verlag, Berlin, 2001); arXiv:physics/0004060 (2000).

Srinivasan et al 1991: Range perception through apparent image speed in freely flying honeybees. MV Srinivasan, M Lehrer, WH Kirchner & SW Zhang, *Vis Neurosci* **6,** 519–535 (1991).

Srinivasan et al 1996: Honeybee navigation en route to the goal: visual flight control and odometry. MV Srinivasan, S Zhang, M Lehrer & TS Collett, *J Exp Biol* **199,** 237–244 (1996).

Since that formative year of having the office next door to Rob de Ruyter van Steveninck when I was a postdoc in Groningen, the fly visual system has seemed to me an ideal testing ground for physicists' ideas. On the other hand, if you think that brains are interesting because you want to understand your own brain, you might believe that insects are a bit of a side show relative to animals that share more of our brain structures—monkeys, cats, or even mice. There are obvious questions of strategy here, including the fact that (perhaps paradoxically) it can be easier to control the behavior of a primate than the behavior of an insect, creating

opportunities for certain kinds of quantitative experiments. There also are questions about how much universality we should expect. Are there things to be learned about brains in general, or is everything about our brain different from that of "lower" animals? Can careful, quantitative analyses of "simpler" systems sharpen the questions that we ask about bigger brains (even if the answers are different), or does each case present such unique challenges? I think it is fair to say that for several decades there has been a strong consensus of the mainstream neuroscience community that the answers to these questions point away from the study of insect brains. Recently, however, there has been substantial growth in a community of scientists interested in exploiting the tools of modern molecular biology to study the brain, and this group of course is attracted to "model organisms" with well developed methods of genetic manipulation, such as the fruit fly *Drosophila melanogaster* and its close relatives. Thus, the coming years are likely to see a resurgence of interest in insect brains, and this should create more opportunities for physicists. It is early days, but here is a selection of papers that may help you in your explorations.

: Find a selection of *Drosophila* articles that point toward quantitative opportunities.

Seelig et al 2010: Two–photon calcium imaging from head–fixed *Drosophila* during optomotor walking behavior. JD Seelig, ME Chiappe, GK Lott, A Dutta, JE Osborne, MB Reiser & V Jayaraman, *Nature Methods* **7,** 535–540 (2010).

The rather astonishing results in Fig 67 are from Griffiths & Tenenbaum (2006). The original work on optimal cue combination was by Ernst & Banks (2002) [cite follow ups!]. [More: Maloney, Wolpert, ... Finally, ambiguous percepts, multistabiity, connections to conscious experience ...]

Bialek & DeWeese 1995: Random switching and optimal processing in the perception of ambiguous signals. W Bialek & M DeWeese, *Phys Rev Lett* **74**, 3077–3080 (1995).

Ernst & Banks 2002: Humans integrate visual and haptic information in a statistically optimal fashion. MO Ernst & MS Banks, *Nature* **415,** 429–433 (2002).

Griffiths & Tennenbaum 2006: Optimal predictions in everyday cognition. TL Griffiths & JB Tennenbaum, *Psychological Science* **17,** 767–773 (2006).

D. Proofreading and active noise reduction

Fluctuations are an essential part of being at thermal equilibrium. Thus, the fact that life operates in a relatively narrow range of temperatures around 300 K means that some level of noise is inevitable. But being alive certainly is not being at thermal equilibrium. Can organisms use their non–equilibrium state to reduce the impact of nominally thermal noise? More generally, can we understand how to take a system in contact with an environment at temperature T, and expend energy, driving it away from equilibrium, in such a way as to reduce the effects of noise?

In his classic lectures *What is Life?*, Schrödinger waxed eloquent about the fidelity with which genetic information is passed from generation to the next, conjuring the image of a gallery with portraits of the Hapsburgs, their oddly shaped lips reproduced across centuries of descendants. Schrödinger was much impressed by the work of Timoféef–Ressovsky, Zimmer and Delbrück, who had determined the cross–section for ionizing radiation to generate mutations, and used this to argue that genes were of the dimensions of single molecules. Thus, the extreme stability of our genetic inheritance could not be based on averaging over many molecules, as a "naive classical physicist" might have thought. Now is a good time to set aside our modern insouciance and allow our ourselves to be astonished, as Schrödinger was, that so many of the phenomena of life are the macroscopic consequences of individual molecular events.

We now teach high school students that the key to the transmission of genetic information is the pairing of bases along the double helix—A pairs with T, C pairs with G, as in Fig 68. This, of course, is the triumph of Watson and Crick's theory of DNA structure.[52] The ideas of templates and structural complementarity which are at the heart of the double helix reappear many times—every time, in fact, that the organism needs to make reliable choices about which molecules to synthesize. But does structural complementarity solve the problem of reliabil-

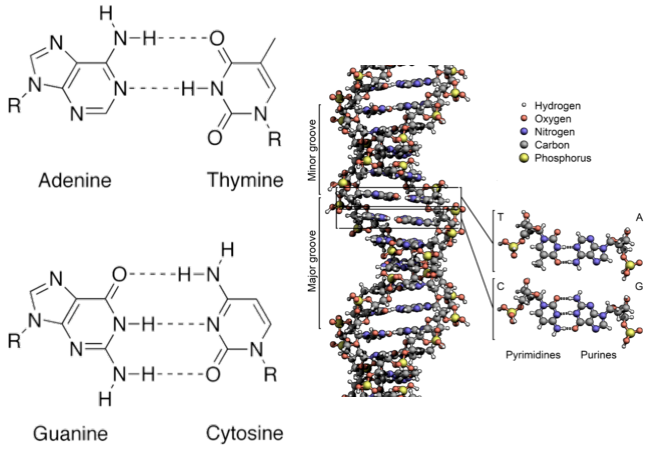

FIG. 68 Base pairing in the Watson–crick structure of DNA [this just grabbed from Wikipedia; need to decide exactly what to show, and redraw]. At left, we see the hydrogen bonding between bases in the correct pairings, showing how they "fit" to satisfy the opportunities for hydrogen bonding, producing structures that are the same width and hence can fit into the double helix, as shown at right. "R" denotes the sugar and phosphate groups, identical for all bases, which form the outer backbone(s) of the helix.

[52] It would be almost silly to think you know something about "biophysics" (whatever you think the word means!) and not understand the interplay of theory and experiment that led to this revolution in the middle of the twentieth century. For a brief tour, see Appendix A.5.

ity in biosynthesis?

The fact that A pairs with T is really the statement that the (free) energy of a correct AT pair is much lower than that of an incorrect AC or AG pair. We should recall the energy scales for chemical bonding. A genuine covalent bond, such as the carbon–carbon or carbon–nitrogen bonds in the interior of the bases, results from the sharing of electrons between the atoms, and the energies are therefore on the scale of several electron volts.[53] Making the wrong base pairs wouldn't require us to break any covalent bonds, so the energy cost will not be this large. If we tried to make an AG pair, it would be so big that it wouldn't fit inside the backbone of the double helix; more precisely, we would have make large distortions of the covalent bonds, and since these are stiff, the energy cost would be very large. On the other hand, if we try to make a CT pair, the backbone will hold the bases so far apart that they can't form hydrogen bonds. Thus, the minimal energy for a "wrong" base pair is the energy of two missing hydrogen bonds, and this is on the order of $10\,k_BT$.

An energy difference of $\Delta F \sim 10\,k_BT$ means that the probability of an *incorrect* base pairing should be, according to the Boltzmann distribution, $e^{-\Delta F/k_BT} \sim 10^{-4}$. A typical protein is three hundred amino acids long, which means that is encoded by nearly one thousand bases; if the error probability is 10^{-4}, then replication of the DNA would introduce roughly one mutation in every tenth protein. For humans, with a billion base pairs in the genome, every child would be born with hundreds of thousands of bases different from his or her parents. If these predicted error rates seem large, they are—real error rates in DNA replication vary across organisms, but are in the range of $10^{-8} - 10^{-12}$, so that entire genomes can be copied almost without any mistakes.

The discrepancy between Boltzmann probabilities and observed error rates is much more widespread. When information encoded in the DNA is read out to make proteins, there are several steps where errors can occur. First is the synthesis of mRNA from the DNA template, a process not unlike the replication of the DNA itself. The "codebook" for translating from the language of bases along the mRNA into amino acids is embodied in the tRNA molecules, which at one end have a triplet of bases (the anti–codon) that is complementary to a particular triplet of bases along the mRNA (the codon), and at their other end is the amino acid that the codon represents. To make such molecules, there are specialized enzymes that recognize the 'bare' tRNA and choose out of the cellular soup the correct amino acid with which to 'charge' the molecule. [the discussion of tRNA and charging could use some sketches!] But some amino acids differ simply by the replacement of a CH_3 group with an H; it we imagine the enzyme recognizing the first amino acid with a binding pocket that is complementary to the CH_3 group, then the second amino acid will also fit, and the binding energy will be weaker only by the loss of non–covalent contacts with the methyl group; it is difficult to see how this could be much more than $\sim 5\,k_BT$, corresponding to error rates $\sim 10^{-2}$. If the error rates in tRNA charging were typically 10^{-2}, almost all proteins would have at least one wrong amino acid; in fact error rates are more like 10^{-4}, so that most proteins have no errors. There is one more step, at the ribosome, where tRNA molecules bind to their complementary sites along the mRNA and the amino acids which they carry are stitched together into proteins, and here too there is a discrepancy between thermodynamics and the observed error probabilities.

Each of the events we have outlined—DNA replication, mRNA synthesis, tRNA charging, and protein synthesis on the ribosome—has its own bewildering array of biochemical details, and is the subject of its own vast literature. As physicists we search for common theoretical principles that can organize this biological complexity, and I think that this problem of accuracy beyond the thermodynamic limit provides a wonderful model for this search. The key ideas go back to Hopfield and Ninio in the 1970s. Their classic papers usually are remembered for having contributed to the solution of the problem of accuracy, a solution termed 'kinetic proofreading,' which we will explore in a moment. But I think they should also be remembered for having recognized that there is a common physics problem that runs through this broad range of different biochemical processes.

To understand the essence of kinetic proofreading, it is useful to recall the problem of Maxwell's demon. Imagine a container partitioned into two chambers by a wall, with a small door in the wall. [again, a sketch would help!] Maxwell conjured the image of a small demon who controls the door. If he[54] see a molecule coming from the right at high speed, he opens the door and allows it to go into the left chamber. Conversely, if he sees a molecule drifting slowly from the left, he opens the door and allows it to enter the right chamber. After some time, all the slow molecules are on the right, all the fast molecules are on the left. But, since the average kinetic energy of the molecules in a gas is proportional to the temperature, the demon has created a temperature difference, hot on the left, cold on the right. This temperature difference can be used to do useful work (e.g., running a heat engine), and thus the demon appears to have created something out

[53] Chemists prefer to think per mole rather than per molecule, and they prefer joules to electron Volts (I won't speak of calories). To have some numbers at your fingertips, remember that at room temperature, $k_BT = 1/40\,\mathrm{eV} = 2.5\,\mathrm{kJ/mole}$.

[54] Why is it obvious that the demon is male?

of nothing, violating the second law of thermodynamics.

There is nothing special about the demon's choice of molecular speed as the criterion for opening the door. It is a simple choice, because the result is a temperature difference, and we can imagine all sorts of appropriately nineteenth century methods for extracting useful work from temperature differences. But if there are two kinds of molecules, A and B, and the demon arranges for the A molecules to accumulate in the left chamber and B molecules to accumulate in the right chamber, then there will be differences in chemical potential between the two chambers, and there must be some way of using this to do work even if as physicists we don't know enough chemistry to figure it out.

Problem 77: Pushing away from equilibrium. Consider a polymer made from A and B monomers. Suppose we start with pure poly–A, and use this as a template to construct a new polymer, much as in DNA replication (but simpler!). Template directed synthesis works because the $A-A$ bond is stronger than the $A-B$ bond by some free energy difference ΔG; we'll use the convention that $\Delta G > 0$. Then if we make a polymer of length N in which a fraction f of the monomers are incorrectly made to be B rather than A, the free energy of the system will have a contribution $Nf\Delta G$ relative to the perfectly copied poly–A. If the errors are made at random, however, then there is a contribution to the entropy of the polymer that comes from the sequence heterogeneity.

(a.) Evaluate the entropy that comes from the random substitutions of A by B. What assumptions are you making in this calculation? Can you imagine these being violated by real molecules?

(b.) Combine the entropy from [a.] with the "bonding" free energy $Nf\Delta G$ to give the total free energy of the polymer. Show that this is minimized at $f_{\rm eq} \propto \exp(-\Delta G/k_BT)$, as expected.

(c.) How much free energy is stored in the polymer when $f < f_{\rm eq}$? Can you give simple expressions when the difference $f_{\rm eq} - f$ is small? What happens if (as we will see below) $f \approx f_{\rm eq}^2$?

The demon's sin is to have generated a state of reduced entropy. We know that to enforce the second law, this non–equilibrium state must be 'paid for' with enough energy to balance the books—to avoid building a perpetual motion machine, the demon must have dissipated an amount of energy equal to or greater than the amount of useful work that can be extracted from his reduction in the entropy of the system. The key insight of Hopfield and Ninio was that the problem of accuracy or low error rates was of this same kind: achieving low error rates, sorting molecular components with a precision beyond that predicted by the Boltzmann distribution, means that the cell is building and maintaining a non–equilibrium state, and it must spend energy in order to do this. Somewhere in the complexity of the biochemistry of these processes there must be steps which dissipate energy, and this has to be harnessed to improve the accuracy of synthesis.

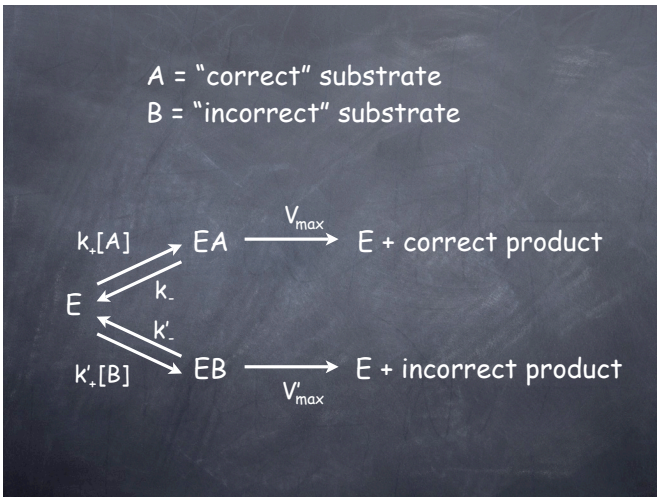

FIG. 69 The simplest kinetic scheme in which an enzyme can choose correct or incorrect molecules out of solution, making correct or incorrect products.

To see how this might work, let's look at the simplest model of a biochemical process catalyzed by an enzyme, as in Fig 69. In essence, the chemical reaction of interest involves choosing among two (or more) substrate molecules, for example the correct and incorrect base at a particular point along the strand of DNA that the cell is trying to replicate or transcribe into mRNA. In order to complete the reaction, the substrate has to bind to the enzyme, and this enzyme–substrate complex can be converted into the product; in order to have any possibility of correcting errors, it must be possible for the substrate to unbind from the enzyme before the conversion to product. With only this minimum number of steps, the kinetics are described by

$$\frac{d[EA]}{dt} = k_+[A][E] - (k_- + V_{\rm max})[EA] \quad (436)$$

$$\frac{d[EB]}{dt} = k'_+[B][E] - (k'_- + V'_{\rm max})[EB] \quad (437)$$

$$[E]_{\rm total} = [EA] + [EB] + [E], \quad (438)$$

where A is the correct substrate, B is the incorrect substrate, and $[E]_{\rm total}$ is the (fixed) total concentration of enzyme molecules. The rate at which correct products are made is given by $V_{\rm max}[EA]$, and the rate of making incorrect products is $V'_{\rm max}[EB]$. If the overall rate of reactions is slow enough not to deplete the substrates (and the cell typically is working hard to make sure this is true!), then we can compute these rates in the steady state approximation.

To compute the rate of errors we don't even need to solve the entire problem. From Eq (436) we can see that, in steady state,

$$[EA] = [E]\frac{k_+[A]}{k_- + V_{\rm max}}; \quad (439)$$

similarly, from Eq (437),

$$[EB] = [E]\frac{k'_+[B]}{k'_- + V'_{\rm max}}. \tag{440}$$

Thus the error probability, or relative rate at which incorrect products are made, is given by

$$f \equiv \frac{\text{rate of making incorrect product}}{\text{rate of making correct product}} \tag{441}$$

$$= \frac{V'_{\rm max}[EB]}{V_{\rm max}[EA]} \tag{442}$$

$$= \left[\frac{k'_+[B]}{k'_- + V'_{\rm max}}\right] \times \left[\frac{k_+[A]}{k_- + V_{\rm max}}\right]^{-1} \times \left[\frac{V'_{\rm max}}{V_{\rm max}}\right] \tag{443}$$

To go further it is useful to notice that all the reactions we are thinking about share one important feature: the actual making and breaking of covalent bonds occurs on 'the other side' of the molecule from the structure that defines correct vs. incorrect [definitely needs a sketch!]. In the case of DNA replication, for example, correctness has to do with the pattern of hydrogen bonding between the bases, on the inside of the helix, while the actual reaction required to incorporate one base into the growing polymer involves the phosphate backbone on the outside of the helix. This makes it unlikely that the rate at which these bonds are formed is sensitive to the correctness of the substrate. Correspondingly, in the cases of interest, it is likely that $V'_{\rm max} \approx V_{\rm max}$, so this is not a source of selectivity. More importantly, from Eq (443) it is clear that, under these conditions, the error probability is minimized if the catalytic rate $V_{\rm max}$ is slow compared with the unbinding rates k_-, k'_-. This makes sense: if the catalytic step itself has no selectivity, then to maximize selectivity one must give the wrong substrate a chance to fall off.

So, when the dust settles, in this simplest kinetic scheme we have shown that the error probability is bounded,

$$f > \left(\frac{k'_+[B]}{k'_-}\right) \Big/ \left(\frac{k_+[A]}{k_-}\right). \tag{444}$$

But this combination of rates and concentrations is exactly what determines the equilibrium binding of A vs B to the enzyme, and hence can be written in terms of thermodynamic quantities,

$$f > \exp\left(-\frac{F_A - F_B}{k_B T}\right), \tag{445}$$

where F_A is the free energy for taking a single molecule of A out of solution and binding to the enzyme, and similarly for B; here binding energies are positive, larger for tighter binding. Thus, we are back where we started, with an error probability determined by the Boltzmann distribution!

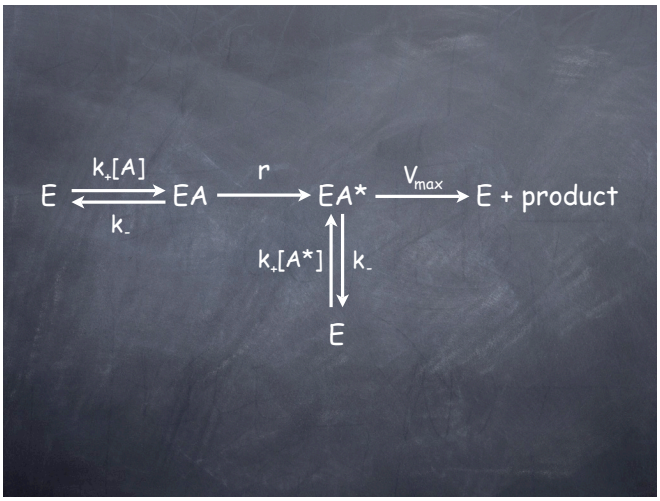

FIG. 70 The simplest scheme for "kinetic proofreading." As described in the text, the key step is an irreversible transition from EA to EA^*, which gives a true second chance for equilibration with the free A molecules.

But the Michaelis–Menten scheme has a natural generalization. Suppose that, after binding, there is an irreversible transition to a new state, at a rate r, and that in this state the substrate can again be released from the enzyme, as in Fig 70. In the simplest case, the events which determine binding and release of the (perhaps modified) substrate are the same as in the initial step, with the same rates. We can carry through the analysis of this kinetic scheme as before, and with the same assumption that catalytic steps ($V_{\rm max}$ and r) have no selectivity, find that

$$f > \exp\left(-\frac{F_A - F_B}{k_B T}\right) \exp\left(-\frac{F_{A^*} - F_{B^*}}{k_B T}\right). \tag{446}$$

But if the molecular interactions that select A over B are the same for A^* vs B^*, we expect $F_{A^*} - F_{B^*} \approx F_A - F_B$, and hence

$$f \to \left[\exp\left(-\frac{F_A - F_B}{k_B T}\right)\right]^2. \tag{447}$$

This is the essence of kinetic proofreading: by introducing an irreversible step into the kinetic scheme, a step which necessarily dissipates energy, it is possible to use the equilibrium selectivity twice, and achieve an error probability which is the square of the nominal limit set by the Boltzmann distribution.

Problem 78: More on the basics of kinetic proofreading. To begin, give the details needed to derive Eq (446). An even better exercise is to go through Hopfield's original paper (Hopfield 1974), pen in hand, filling in all the missing steps. Then consider the following:

(a.) In the simplest scheme, we saw that maximum selectivity occurs when $V_{\max}$ is slow compared with k_-. Is there a similar condition in the proofreading scheme? What does this tell us about the progress of the enzymatic cycle? More specifically, what is the fate of the typical substrate which binds to the enzyme? Is it converted to product, or ejected as A^*?

(b.) Consider a generalization of the kinetic scheme in Fig 70 such that the nominally irreversible step with rate r is in fact reversible, with the reverse reaction at rate r'. To be general, imagine also the binding and unbinding of A^* can occur with rates that are different from the rates for A. Now there are detailed balance conditions that connect these different rates. Write down these conditions, and show how they effect the error probability. Can you say something general here? In particular, can you show how these conditions enforce the Boltzmann error rate in the absence of energy dissipation, no matter how many times the enzyme 'looks' at the substrate?

How does this general idea of proofreading connect with the real biochemistry of these systems? In some sense the case of DNA replication (or transcription) is most obvious, as shown in Fig 71. All of the nucleotides which are incorporated into the growing strands of DNA or RNA start as nucleotide triphosphates, but once the final structure is formed only one phosphate is part of the backbone. Thus, at some point in the process, the 'high energy' phosphate bond must be cleaved, releasing roughly $20k_BT$ of free energy. If this is the irreversible step, then is must be possible for the enzyme which catalyzes the growth of the polymer to release the nucleotide after this cleavage, which means after it has been attached to the backbone of the growing chain. Thus, to proofread, the enzyme must be not only a 'polymerase' (catalyzing the polymerization reaction) it must also be an 'exonuclease' (catalyzing the removal of nucleotides from the polymer). It had been known almost since the discovery of the polymerase that it also had exonuclease activity, but it took the idea of kinetic proofreading to explain how this was connected, through energy dissipation, to proofreading and error correction. In the charging of tRNA, the process actually starts with an ATP molecule being cleaved, leaving an AMP attached to the amino acid before it reacts with the tRNA. In protein synthesis, the sequence of reactions is much more complex, but again there is an obligatory cleavage of a nucleotide triphosphate (in this case GTP → GDP). All of these examples are qualitatively consistent with the proofreading scenario,[55] and especially in the case of tRNA charging it has been possible to pursue a more quantitative connection between theory and experiment [do we want to say more about this?].

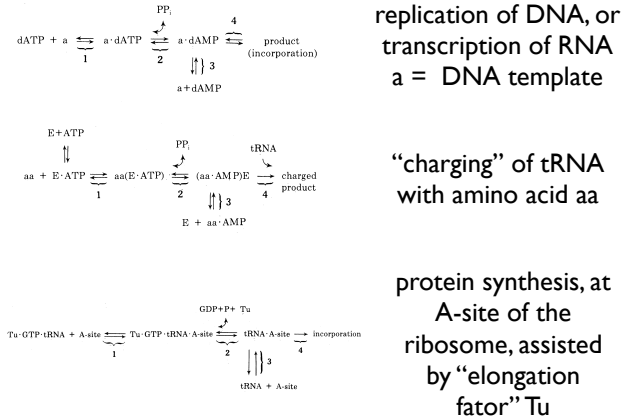

FIG. 71 Connecting the proofreading scheme to specific biochemical process, from Hopfield (1974). At the top, nucleotide triphosphates are incorporated as monophosphates in DNA replication or the transcription to mRNA. In the middle panel, the charging of tRNA molecules with amino acids, involving an extra ATP. At bottom, a very simplified view of protein synthesis, in which the GTP/exchange by the protein Tu provides the energy for proofreading at the ribosome.

Kinetic proofreading not only solves a fundamental problem—the problem which Schrödinger confronted in the Hapsburg portraits—it also has been a source of new questions and ideas. If the accuracy of DNA replication depends not only on intrinsic properties of the DNA but also on the detailed kinetics of the enzymes involved in replication, then the rate of mutations itself can be changed by mutations. It has long been known that there are 'mutator strains' of bacteria which have unusually high error rates, and we now know that that these strains simply have aspects of the proofreading apparatus disabled. One could imagine subtler changes, so that the mutation rate would become a quantitative trait; in this case the dynamics of evolution would be very different, since fluctuations along one "direction" in the space of genomes would change the rate of movement along all directions. Also, since accuracy depends on energy dissipation, in an environment with limited nutrients there is a tradeoff between the speed of growth and the fidelity with which genetic information is passed to the next generations; there is an optimization problem to be solved here, and ... [say something definite re Kurland, Ehrenberg, ... maybe have a problem?]. In protein synthesis, accuracy and even the overall kinetics will be affected by the availability of the different charged tRNAs, and this is under physiological control, so again there is the pos-

[55] Hopfield has also emphasized that there are kinetic schemes in which proofreading still proceeds through energy dissipating steps, but if the enzymes have some memory for past events then the synthesis and dissipation can be separated in time, erasing some of the more obvious signatures from the simpler scheme. This may be especially important in thinking about more complex examples, such as protein synthesis on the ribosome or DNA replication in higher eukaryotes. [is there a good problem to give here?]

sibility that, especially for fast growing bacteria where the problems are most serious, there is some tuning or optimization to be done.

Problem 79: Controlling the pace of evolution? [take the students through a simple version of Magnasco & Thaler. Introduces ideas of evolutionary landscape, connect back to discussion of reaction rates ...]

Problem 80: Optimizing tRNA pools. There is a separate tRNA complementary to each of the 60 codons which code for amino acids (the remaining four codons stand for 'start' and 'stop'). The frequency with which these codons are used in the genome varies widely, both because proteins do not use all 20 amino acids equally and because different organisms use different synonymous codons (that is, those which code for the same amino acid) with different frequencies. But, when it comes time to make protein, the cell needs access to the appropriate population of charged tRNAs. Naively one might expect that, if the supply of tRNA is limiting the rate at which a bacterium can make proteins and grow, then it would be good to have a supply of tRNA in proportion to how often the corresponding codon gets used. Let's see if this is right. Suppose that protein synthesis is limited by arrival of the tRNA at the ribosome. Then the time required to incorporate one amino acid coded by codon i is $t_i \sim 1/k[tRNA_i]$, where k is a second order rate constant.

(a.) [try to sort out how rate of ribosome turnover compares with diffusion limited rate of arrival of tRNAs]

(b.) The average time required to incorporate one amino acid is $\bar{t} = \sum_i p_i/k[tRNA_i]$, where p_i is the probability of codon i appearing in the cell's mRNA. If the cell can only afford a limited amount of tRNA, the natural constraint is on the total $\sum_i [tRNA_i]$. How should the individual concentrations be arranged to minimize the mean incorporation time $\bar{t}$? Is this surprising?

(c.) You might be tempted to say that, if the goal is to synthesize proteins as rapidly as possible, and the rates are limited by the arrival of tRNAs, then we should maximize the mean rate, $\sum_i p_i k[tRNA_i]$. Why is this wrong?

The ideas of kinetic proofreading may be even more generally applicable than envisioned by Hopfield and Ninio. There are many signal transduction processes that start with a receptor binding event at the cell surface and trigger a cascade of protein phosphorylation reactions;[56] the phosphate groups are pulled from ATP, so phosphorylation is a prototypically irreversible, energy consuming reaction. In the immune system [need a figure here!] it has been suggested that this can provide multiple stages

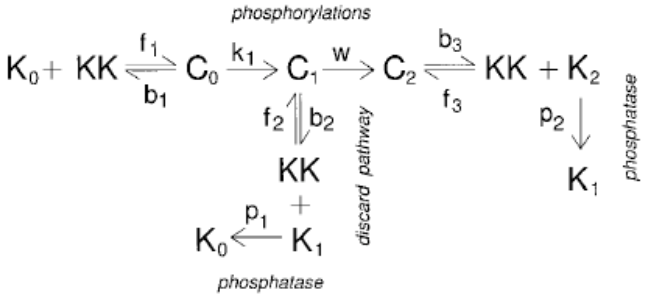

FIG. 72 Kinetic proofreading in the phosphorylation of a kinase (K) by a kinase–kinase (KK), from Swain & Siggia (2002). Activation of the kinase requires two steps of phosphorylation, and in this scheme the the kinase–kinase can dissociate from its substrate after have transferred just one phosphate group. K_0, K_1 and K_2 denote the kinase with zero, one and two attached phosphate groups, respectively.

of proofreading, contributing to self/non–self discrimination. More generally, as shown in Fig 72, if activation of an enzyme requires two steps of phosphorylation, then these steps can be arranged in a proofreading scheme. Because there are many such pathways in the cell, proofreading this case could increase specificity and reduce crosstalk.

Watson and Crick understood that the double helical structure of DNA, with its complementary strands, suggested a mechanism for the copying of genetic information from one generation to the next. But they also realized that the helical structure creates a problem, since the strands are entangled; the problem is most obvious in bacteria, where the chromosomes close into circles, but with very long molecules one couldn't rely on spontaneous untangling even if there is no formal topological obstruction. Eventually it was discovered that there is a remarkable set of enzymes that catalyze changes in the topology of circular DNA molecules, allowing the strands to pass through one another. In the process of relieving entanglement, these "topoisomerases" also reduce the energy stored in the supercoiling of these polymers [should say more about this here—an excuse to talk about link, writhe and twist, etc.; certainly needs a figure]. The problem is that being truly unlinked is a global property of the molecules, while the enzymes act locally. In the simplest models, then, topoisomerases would remove the obstacles to changing topology, but couldn't shift the probability of being unlinked from its equilibrium value. Because making links or knots restricts the entropy of the molecule, there is an equilibrium bias in favor of unlinking, but this seems insufficient for cellular function.

[56] [Should have said something about this already!] Many proteins are activated by the covalent addition of phosphate groups, a reaction termed phosphorylation. Enzymes that catalyze the transfer of phosphate groups are called kinases, and these enzymes often are usually specific for their substrates, whether these are smaller molecules or proteins. Importantly, come kinases themselves are activated by phosphorylation, and the enzymes that carry out this activation step are termed kinase kinases.

Indeed, as shown in Fig 73, topoisomerases seem to leave fewer links than expected from the Boltzmann distribution even in test tube experiments, and if we look at the details of the biochemical steps involved, we can identify a series of steps that are equivalent to proofreading by the topoisomerases [I'd like to explain this better!].

The ideas of proofreading have recently been revitalized by the opportunity to observe, more or less directly, the individual molecular events responsible for error correction. The key to this new generation of experiments is the realization that molecules such as RNA polymerase are "molecular motors" that move along the DNA strand as they function. Each step in this movement is presumably on the scale of the distance between bases along the DNA, $d \sim 3.4\,\mathring{A}$. The energy to drive this motion comes from breaking the phosphate bonds of the input nucleotides, and is on the scale of $\sim 10 k_B T$. Thus the forces involved are $F \sim 10 k_B T/d \sim 100\,\mathrm{pN}$.

When a dielectric sphere sits in an electric field, it polarizes, and the direction of the polarization is such that it lowers the energy. This means that the energy of the sphere is lower in regions of high electric field. Since the energy is proportional to the square of the field, this is true even if the field is oscillating in time. In particular, if we focus a light beam in a microscope, then the light intensity is higher in the focus, and light intensity is just the square of the electric field, so we expect that small dielectric objects will be attracted to focal spots, and this

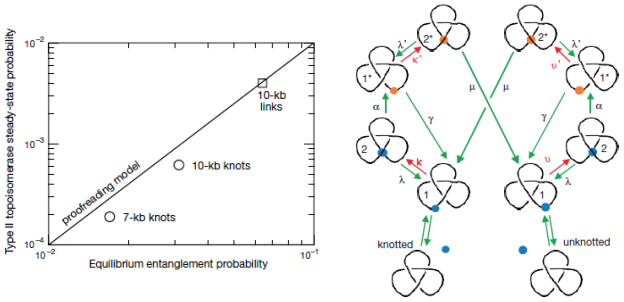

FIG. 73 Kinetic proofreading in DNA unlinking, from Yan et al (1999). At left, experimental results redrawn from Rybenkov et al (1997), showing that topoisomerases reaching a linking probability roughly equal to the square of the expected equilibrium probability, suggesting a proofreading scheme. At right, a kinetic scheme illustrating the possibility of proofreading, Active topoisomerase molecules are shown in red, inactive in blue; green arrows denote transitions that are insensitive to the topology, while all sensitivity is contained in the red arrows. This kinetic scheme is essentially a "folded" version of Hopfield's original Fig 70.

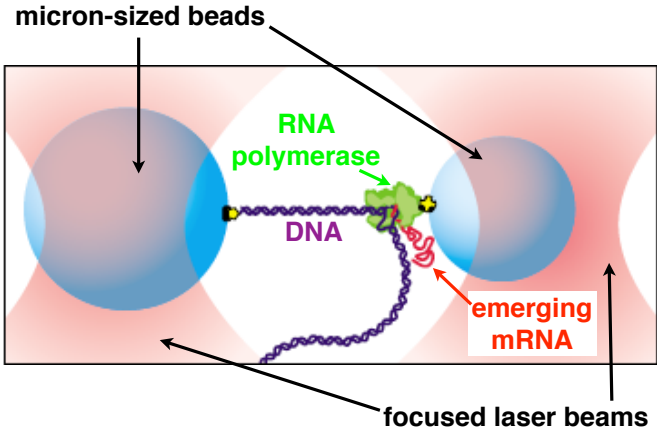

FIG. 74 Schematic of an experiment to observe the function of RNA polymerase with single base–pair resolution, from Shaevitz et al (2003). A laser beam is split, and the two resulting beams are focused to make "optical traps" for two micron–sized beads. Attached to one bead is a double stranded DNA molecule, and attached to the other is an RNA polymerase molecule. As the polymerase synthesizes mRNA, it "walks" along the DNA and the tether between the two beads is shortened. The intensities of the two beams are set so that the left hand trap was stiffer, insuring that most of the motion appears as a displacement of the right hand bead, which is measured by projecting scattered light onto a position–sensitive detector.

is called "optical trapping." Importantly, with realistic light intensities, the forces on micron–sized particles as they move in an optical trap indeed are on the scale of picoNewtons, so it is possible to "hold" a molecular motor in place.

Problem 81: Optical trapping. The key to the experiments here is the fact that small, neutral particles can be trapped at the focus of a laser beam, and that the forces generated in this way are on the same scale as those generated by individual biological motor molecules, such as the RNA polymerase. Take the students through this!

In Figure 74 we see the schematic of an optical trapping experiment on the RNA polymerase. Successive generations of technical improvements in these experiments have made it possible to track the motion of the polymerase with a resolution fine enough to see it "step" from one base pair to the next, as in Fig 75. Importantly, in these experiments one can bathe the sample in a solution containing different nucleotides. If we add ITP, which is not one of the standard four bases, it will sometimes be incorporated into the growing mRNA strand, but this is

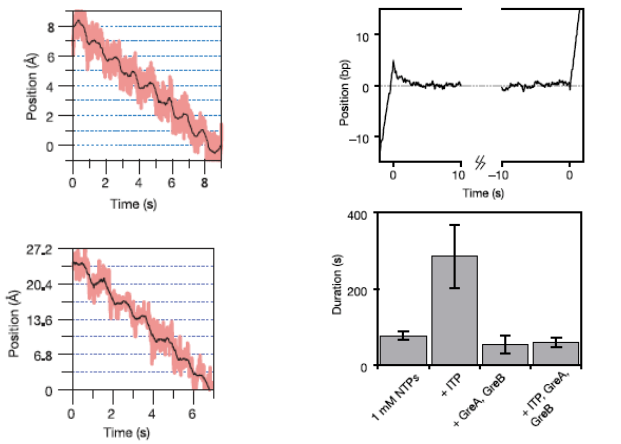

FIG. 75 Motion of the RNA polymerase along DNA. At left, from Abbondanzieri et al (2005). Top, the position of the right hand bead from Fig 74 as the trap is moved in 1 Å steps, to show that these can be resolved. Bottom, the active motion of the bead as the RNA polymerase synthesizes rRNA, showing the expected steps of 3.4 Å. [are the black lines median filtering?] At right, from Shaevitz et al (2005). Top, the average trajectory of the RNA polymerase aligned on the start and end of long pauses. Bottom, the mean duration of pauses under different conditions, notably the addition of the "wrong" nucleotide ITP.

always a mistake. Under these conditions we can observe an increased frequency of "pauses" in the motion of the polymerase, followed by backtracking of 1–10 base pairs along a relatively stereotyped trajectory. If we remove from RNA polymerase the subunits thought to be involved in proofreading, then these error–induced pauses become very long.

[Need a summary on kinetic proofreading, segue to active filtering]

There is another broad class of examples in which there seems to be a discrepancy between the noise expected at thermal equilibrium and the performance of biological systems, and this is in the measurement of small displacements. In our inner ear, and in the ears of all other vertebrate animals, motions are sensed by "hair cells," so named because of the tuft of "hairs" (more properly, stereocilia) that project from their top surface as in Fig 76. Although we usually think of ears as responding to airborne sounds, in fact there are multiple chambers in the ear, some of which respond to sound, and others of which respond to lower frequency motions generated by rotation of our head, the largely constant force of gravity or ground borne vibrations. The core of all these systems, however, is the hair cell. When the stereocilia are bent, channels in the cell membrane open and close, and this modulates an ionic current, as in other receptor cells that we have seen before. In a variety of systems it has been possible to open these organs, or even dissect out the hair cells, and to make direct mechanical measurements on the stereocilia. Typically, the bundle of hairs moves as a unit, and the stiffness is in the range of $\kappa \sim 10^{-3}\,\mathrm{N/m}$ or less. This implies that the Brownian motion of the bundle should have an amplitude $\delta x_{\rm rms} = \sqrt{k_B T/\kappa} \sim 2\,\mathrm{nm}$. This seems small (remember that the stereocilia have lengths measured in microns), but ... .

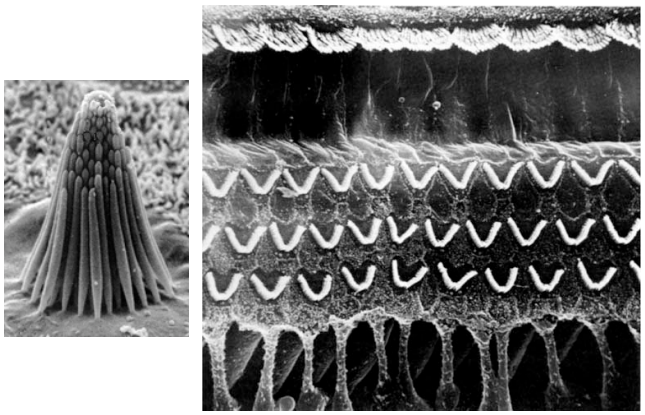

FIG. 76 Hair cells of the vertebrate inner ear [find better images, with scale bars!]. At left, in the bullfrog sacculus, from http://www.hhmi.org/senses/c120.html. At right, in the mammalian cochlea, three rows of "outer" hair cells and one row of "inner" hair cells at top, from Dallos (1984).

There is a particular species of neotropical frog, for example, that exhibits clear behavioral responses to vibrations of the ground that have an amplitude of ~ 1 Å. Individual neurons which carry signals from the hair cells in the sacculus the to brain actually *saturate* in response to vibrations of just ~ 10 Å $= 1$ nm. Although there are controversies about the precise numbers, the motions of our eardrum in response to sounds we can barely hear are similarly on the atomic scale. Invertebrates don't use hair cells, but they also have mechanical sensors, and many of these too respond reliably to motions in the Ångström or even sub–Ångström range.

By itself, the order–of–magnitude (or more) discrepancy between the amplitude of Brownian motion and the threshold of sensation might or might not be a problem (we'll come back to this). But surely it motivates us to ask if, by analogy with kinetic proofreading, it is possible to lower the effective noise level by pushing the system away from thermal equilibrium. This also is an interesting physics problem, independent of its connection to biology.

Consider a mass hanging from a spring, subject to drag as it moves through the surrounding fluid, as in Fig 77. By itself, the dynamics of this system are described by

the Langevin equation, [point back!]

$$m\frac{d^2x(t)}{dt^2} + \gamma\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) + \zeta(t), \qquad (448)$$

where $F_{\rm ext}$ denotes external forces acting on the system and the Langevin force obeys

$$\langle \zeta(t)\zeta F(t')\rangle = 2\gamma k_B T\delta(t-t'). \qquad (449)$$

But suppose that we measure the position of the mass, differentiate to obtain the velocity, and then apply a "feedback" force proportional to this velocity, $F_{\rm feedback} = -\eta dx(t)/dt$; then we have

$$m\frac{d^2x(t)}{dt^2} + \gamma\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) + \zeta(t) + F_{\rm feedback}(t) \qquad (450)$$

$$= F_{\rm ext}(t) + \zeta(t) - \eta\frac{dx(t)}{dt} \qquad (451)$$

$$m\frac{d^2x(t)}{dt^2} + (\gamma+\eta)\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) + \zeta(t). \qquad (452)$$

This system is equivalent to one with a new drag coefficient $\gamma' = \gamma + \eta$. But the fluctuating force hasn't changed—the molecules of the fluid don't know that we are applying feedback—so we can write

$$m\frac{d^2x(t)}{dt^2} + \gamma'\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) + \zeta(t) \qquad (453)$$

$$\langle \zeta(t)\zeta F(t')\rangle = 2\gamma k_B T\delta(t-t') = 2\gamma' k_B T_{\rm eff}\delta(t-t'), \qquad (454)$$

where $T_{\rm eff} = T\gamma/\gamma' = T\gamma/(\gamma+\eta)$. Thus, by observing the system and applying a feedback force, we synthesize a system which is, effectively, colder and thus has (in some obvious sense, but we will need to be careful) less thermal noise.

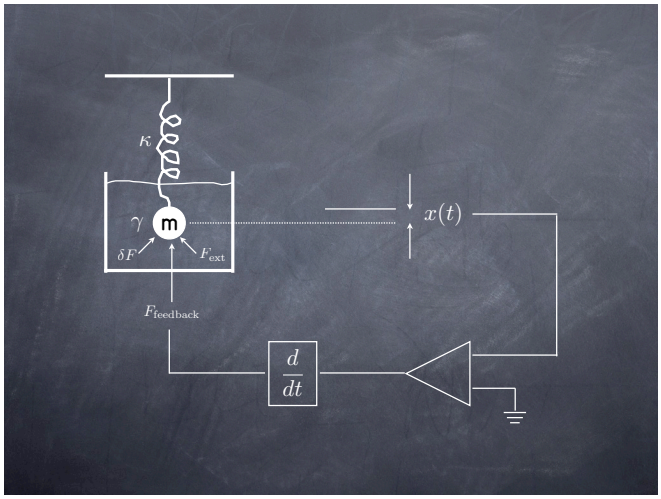

FIG. 77 A schematic of active feedback, in which we observe the position of a mass on a spring and apply a force proportional to the velocity. This can serve to enhance or compensate the intrinsic drag γ, but since it is generated by an active mechanism (symbolically, through the amplifer) there need not be an associated change in the magnitude of the Langevin force, as there would be at thermal equilibrium.

This idea of "active cooling" is very old, but it has received new attention in the attempt to build very sensitive displacement detectors, e.g. for the detection of gravitational waves. A recent example placed a one gram mass in a laser interferometer and used the change in radiation pressure on the mass as function of its position to generate the feedback force; this is different in detail from the model above, but similar in spirit. The result was that the effective temperature could be brought down from $\sim 300\,{\rm K}$ to $\sim 7\times10^{-3}\,{\rm K}$, a reduction of roughly $40,000\times$, and this seems to be limited by noise in the laser itself.

It is important to be clear about exactly which measures of noise are reduced, and which are not. The mean–square displacement of the oscillator—and hence, by equipartition, the apparent temperature—has been reduced. But when we try to drive the system with a force at the resonant frequency, the added damping means that it is more resistant, and hence the response to a given force is smaller. Thus if we ask for the minimum force that we must apply (on resonance) to displace the oscillator by one standard deviation, this threshold force actually goes up, as if the system had more noise, not less. Finally, if we imagine that we can observe the position of the oscillator over a very long time, then what matters for detecting a small applied force at the resonant frequency is the spectral density of force noise, and this hasn't changed at all.

Problem 82: Effective noise levels. Do the real calculations required to verify the statements in the previous paragraph. These are not difficult.

As alternative to actively damping the oscillator, we can try to actively *undamp*, using feedback of opposite sign:

$$m\frac{d^2x(t)}{dt^2} + \gamma\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) + \zeta(t) + F_{\rm feedback}(t) \tag{455}$$

$$= F_{\rm ext}(t) + \zeta(t) + \eta\frac{dx(t)}{dt} \tag{456}$$

$$m\frac{d^2x(t)}{dt^2} + (\gamma-\eta)\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) + \zeta(t). \tag{457}$$

Now the variance of the displacement is larger,

$$\langle(\delta x)^2\rangle = \frac{k_BT_{\rm eff}}{\kappa} = \frac{k_BT}{\kappa}\cdot\frac{\gamma}{\gamma-\eta}, \tag{458}$$

but the sensitivity to forces applied on resonance is also enhanced. If we have $F_{\rm ext}(t) = F_0\cos(\omega_0 t)$, with $\omega_0 = \sqrt{\kappa/m}$, then the displacement will be $x(t) = x_0\sin(\omega_0 t)$, with $x_0 = F_0/[(\gamma-\eta)\omega_0]$. Thus the signal–to–noise ratio in a snapshot of the motion becomes

$$\frac{x_0^2}{\langle(\delta x)^2\rangle} = \frac{F_0^2}{(\gamma-\eta)^2\omega_0^2}\cdot\frac{\gamma-\eta}{\gamma}\frac{\kappa}{k_BT} = \left[\frac{\kappa F_0^2}{(\gamma\omega_0^2)^2}\frac{1}{k_BT}\right]\cdot\frac{\gamma}{\gamma-\eta}. \tag{459}$$

Thus, in this case the signal–to–noise ratio for a snapshot of the position goes up in proportion to the amount of active 'undamping.'

We can understand the impact of active undamping as a narrowing of the system bandwidth, or a sharpening of the resonance around ω_0. Both the external force and the Langevin force drive the system in the same way. The difference is that we are considering an external force at the resonant frequency, while the Langevin force is white noise, with equal power at all frequencies. By sharpening the resonance, active undamping reduces the total impact of this noise; since the bandwidth of the resonance is proportional to $\gamma-\eta$, the enhancement of the signal–to–noise ratio is also in proportion to this factor.

Taken at face value it seems that we can increase the signal–to–noise ratio by an arbitrarily large factor—if we increase η so that $\gamma-\eta \to 0$, the resonance becomes infinitely sharp and it becomes possible to detect arbitrarily small forces from just an instantaneous look at the position x. Any recipe for detecting arbitrarily small signals should be suspect, but what actually limits the growth of the signal–to–noise ratio in this case?

First, it should be clear that the increased SNR comes at a cost. In a system with a sharp resonance, the time scale for response becomes long in inverse proportion to the bandwidth. Thus, as we let $\gamma-\eta \to 0$, the current position $x(t)$ becomes dependent on the forces $F_{\rm ext}(t)$ in distant past. This is a serious issue, but it doesn't really set a limit to the smallest force we can detect.

Problem 83: A reminder about Green functions. The solution to the equation

$$m\frac{d^2x(t)}{dt^2} + (\gamma-\eta)\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) \tag{460}$$

can be written in the form

$$x(t) = \int dt'\, G(t-t')F_{\rm ext}(t'), \tag{461}$$

where $G(\tau)$ is the Green function or (time domain) linear response function. Find $G(\tau)$, and verify that as $\gamma-\eta \to 0$ this function acquires weight at very large τ, corresponding to a very long memory or strongly nonlocal responses.

A second limit to the signal–to–noise ratio is set by noise in the amplifier itself. This certainly is a practical problem, and there may even be a fundamental problem, since linear amplifiers have a minimum level of noise set by quantum mechanics. There is some very interesting physics here, and (confession time) there was a time when I worked very hard to convince myself that these quantum limits to measurement could be relevant to biological systems. This project failed, and I would rather not revisit old failures, so let's skip this one.

The third consideration which limits the narrowing of the bandwidth is the finite power output of any real amplifier. As we let $\gamma - \eta \to 0$, the amplitude of motion in response to a force at resonance grows as $1/(\gamma - \eta)$, and since there is a real drag force $-\gamma(dx/dt)$ the amplifier must dissipate power to drive these ever larger motions. At some point this power requirement will become overwhelming, and the simple model $F_{\rm feedback} = +\eta(dx/dt)$ has to break down. Intuitively, we expect that as x becomes larger, the strength of the feedback will decrease, so we can describe at least the beginning of this power limitation we can write

$$\eta \to \eta(x) \approx \eta_0[1 - (x/x_s)^2 + \cdots], \tag{462}$$

where x_s is the scale on which the amplifier loses linearity. Then we have

$$m\frac{d^2x(t)}{dt^2} + (\gamma - \eta_0)\frac{dx(t)}{dt} + \frac{\eta_0}{x_s^2}x^2(t)\frac{dx(t)}{dt} + \kappa x(t) = F_{\rm ext}(t) + \delta F(t). \tag{463}$$

This equation has several important features.

First, $\gamma = \eta_0$ is a bifurcation point. If $\gamma > \eta_0$, then in the absence of forces any small displacement from $x = 0$ will decay with time. In contrast, for $\gamma < \eta_0$, small displacements will oscillate and grow until the nonlinear term $\sim x^2(dx/dt)$ becomes significant. This is an example of a Hopf bifurcation [should we say some more technical things here about the kinds of bifurcations and the defining features of Hopf?]. Second, if we poise the system precisely at the bifurcation point, and drive it with a resonant force, then neglecting noise we have

$$m\frac{d^2x(t)}{dt^2} + \frac{\gamma}{x_s^2}x^2(t)\frac{dx(t)}{dt} + \kappa x(t) = F_0\cos(\omega_0 t). \tag{464}$$

Guessing that the solution is of the form $x(t) \approx x_0\sin(\omega_0 t)$, we note that

$$x^2(t)\frac{dx(t)}{dt} \approx \omega_0 x_0^3 \sin^2(\omega_0 t)\cos(\omega t) \tag{465}$$

$$= \frac{1}{4}\omega_0 x_0^3\left[\cos(\omega_0 t) - \cos(3\omega_0 t)\right]; \tag{466}$$

in the limit that the resonance is sharp, we know that the term at frequency $3\omega_0$ can't really drive the system, so we neglect this. Thus we have

$$\frac{\gamma\omega_0}{4x_s^2}x_0^3 = F_0, \tag{467}$$

or

$$x(t) = \left[\frac{4F_0x_s^2}{\gamma\omega_0}\right]^{1/3}\sin(\omega_0 t). \tag{468}$$

Thus, the response to applied forces is nonanalytic (at least in the absence of noise); the slope of the response at $F_0 = 0$ is infinite, as one expects from the linear equation above, but the response to any finite force is finite.

The fractional power behavior in Eq (468) connects to a well known but very puzzling fact about the auditory system. As with any nonlinear system, if we stimulate the ear with sine waves at frequencies f_1 and f_2, we can hear "combination tones" built out of these fundamentals: $f_1 \pm f_2$, $2f_1 - f_2$, and so on. In the human ear, the term $2f_1 - f_2$ (with $f_1 < f_2$) is especially prominent. What is surprising is that the subjective intensity of this combination tone is proportional to the intensity of the fundamental tones. If we imagine that combination tones arise from a weak nonlinearity that could be treated in perturbation theory, we would predict that if the input tones have amplitudes A_1 and A_2, then the amplitude of the combination tone should be $A_{2f_1-f_2} \propto A_1^2A_2$. In contrast, the model poised precisely at the bifurcation point predicts $A_{2f_1-f_2} \propto (A_1^2A_2)^{1/3}$, so that if we double the intensity of the input sounds we also double the intensity of the combination tone, as observed.

Problem 84: Combination tones. Do honest calculations to verify the statements about combination tones in the previous paragraph. Contrast the predictions far from the bifurcation point, where perturbation theory is applicable, with the predictions at the bifurcation point.

What happens to the nominally infinite signal–to–noise ratio in the linear model? As we increase the feedback η, the mean square displacement increases, but Eq (462) tells us that at larger x the effective strength of the feedback term decreases. We can try to see what will happen by asking for self–consistency. Suppose we replace the x–dependent value of the feedback term by an effective feedback strength which is given by the average,

$$\eta_{\rm eff} \equiv \langle\eta(x)\rangle = \eta_0[1 - \langle x^2\rangle/x_s^2]. \tag{469}$$

But if we have an effective feedback term we can go back to the linear problem, and then Eq (458) tells us that

$$\langle x^2\rangle = \frac{k_BT}{\kappa}\cdot\frac{\gamma}{\gamma - \eta_{\rm eff}}. \tag{470}$$

Combining these equations gives us a self–consistent equation for the position variance $\langle x^2 \rangle$,

$$\frac{\eta_0}{\gamma x_s^2}\langle x^2\rangle^2 + \left(1 - \frac{\eta_0}{\gamma}\right)\langle x^2\rangle = \frac{k_B T}{\kappa}. \quad (471)$$

Even if we let the strength of the bare feedback η_0 become infinitely large, this equation predicts that the effective feedback term will remain finite, and in particular we always have $\eta_{\rm eff} < \gamma$, so we can never cross the bifurcation, at least in this approximation. Concretely, solving Eq (471) and substituting back into Eq (469) for the effective feedback, we find

$$\lim_{\eta_0 \to \infty} \frac{\gamma - \eta_{\rm eff}}{\gamma} = \frac{k_B T}{\kappa x_s^2}. \quad (472)$$

Thus, the system can narrow its bandwidth to an extent that is limited by the dynamic range of the feedback amplifier, which in turn is related to its power output. Since active narrowing of the bandwidth reduces the effective noise level below the expected thermal noise, we have a situation every much analogous to kinetic proofreading: we can do better than Boltzmann, but it costs energy, and the more energy the system expends, the better it can do.

Problem 85: Noise levels in nonlinear feedback. Start by verifying Eq (472). In the same approximation, calculate the response to applied forces, and show that the smallest force which be detected above the noise has been reduced by a factor $\sim \sqrt{\kappa x_s^2/k_B T}$ relative to what we would have without feedack. Then, there are several things to worry about.

(a.) We have given two analyses. In the first, leading to Eq (468), we neglect noise and take the nonlinearities seriously, finding that the response to small forces is non–analytic. In the second, leading to Eq (472) we treat the crucial nonlinear terms a a self–consistently determined linear feedback, and noise is included. In this second approach, the response to applied forces is linear. Can you reconcile these approaches? Presumably the first approach is valid if the applied forces produce displacements much larger than the noise level. Does this mean that the noise serves to "round" the nonanalytic behavior near $F = 0$?

(b.) How do your results in (a.) effect your estimates of the smallest force that can be detected above the noise?

(c.) You might be worried that our self–consistent approximation is a bit crude. An alternative is to simulate Eq (463) numerically, reminding yourself of the discussion in Section II.A about how to treat the Langevin force. Compare the results of your simulation with the predictions of the self–consistent approximation, for example Eq (471).

(d.)You could also try an alternative analytic approach. If we rewrite Eq (463) in the absence of external forces as

$$\frac{dx(t)}{dt} = v(t) \quad (473)$$

$$m\frac{dv(t)}{dt} = -\left[\gamma - \eta_0\left(1 - \frac{x^2(t)}{x_s^2}\right)\right]v(t) - \kappa x(t) + \delta F(t), \quad (474)$$

then you should be able to derive a Fokker–Planck or diffusion–like equation for the probability $P(x, v)$ of finding the system with instantaneous position x and velocity v. Can you find the steady state solution? How does this compare with your numerical results?

What do we learn from all this? Although there are limits, active feedback (with either sign) makes it possible to detect smaller signals than might otherwise be possible given the level of thermal noise. Pushing the system away from equilibrium, we spend energy to improve performance. This sounds like the sort of thing biological systems might exploit.

If thermal noise is important, then it it useful to think about the bandwidth the system is using as it "listens" (in this case, literally) to its input, and the resulting exchange of energy. We recall that in a resonator, the time scale on which oscillations decay away is $\tau \sim 1/\Delta f$, where Δf is the range of frequencies under the resonant peak. Thus if we excite the resonator to an amplitude such that it stores energy E, this energy also decays away on a time scale $\sim \tau$. But in thermal equilibrium we know that the average energy is not zero, but rather $k_B T$, so the surrounding heat bath must provide a flux of power $\sim k_B T/\tau \sim k_B T \Delta f$ to balance the dissipation. If we want to detect incoming signals above the background of thermal noise, then these signals have to deliver a comparable amount of power. A more careful calculation shows that this "thermal noise power" is $P = 4 k_B T \Delta f$.

Problem 86: Acoustic cross–sections and detailed balance. Use idea of thermal noise power to derive limit on absorption cross–section averaged over directions. Emphasize connection to Einstein's argument about A and B coefficients. Maybe look at data on the ear in relation to this limit?

Estimates of the power entering the inner ear at the threshold of hearing are $P \sim 4 \times 10^{-19}\,$W. This suggests that, to be sure the signals are above thermal noise, the ear must operate with a bandwidth of less than $\Delta f \sim 100\,$Hz. There are several ways of seeing that this is about right. If we record the responses of individual neurons emerging from the cochlea of animals like us, and we can see that these responses are tuned. More quantitatively, as in Fig [**], we can measure the sound pressure required to keep the neuron generating spikes at some fixed rate, and see how this varies with the frequency of pure tone inputs. This input required for constant output is minimal at one "characteristic frequency" of the neuron, and rises steeply away from this minimum; for neurons with characteristic frequencies in the range of 1 kHz, the bandwidths are indeed $\Delta f \sim 100\,$Hz. One can

also try to measure the effective bandwidth in human observers, either by asking listeners to detect a tone in a noisy background and seeing how detection performance varies with the width of the noise, or by testing when one tone impairs the detection of another. More recently it has been possible to record the responses from individual receptor cells, as in Fig [This paragraph needs figures with recordings from primary auditory neurons and hair cells; be sure that these are properly referenced at the end of the section.] All of these bandwidth estimates are in rough agreement, and also agree with the estimate based on comparing thermal noise with the power entering the ear at threshold, suggesting that filtering—in addition to its role in decomposing sounds into their constituent tones—really is essential in limiting the impact of noise. It is important that the resonance or filter which defines this bandwidth actually be in a part of the system where it can act to reject the dominant source of thermal noise. For example, if we think of the vibration sensitive frog, placing the frog on a resonant table would mean that the whole system had a narrower bandwidth, but this would do nothing to reduce the impact of random motions of the stereocilia. It is extremely implausible that the passive mechanics of the stereocilia themselves can generate this narrow bandwidth.

Problem 87: Stereocilium mechanics. Use the image of the hair bundle in Fig ** to estimate the mass and drag coefficient of the bundle as it moves through the surrounding fluid, which you can assume is water. Is the system naturally resonant? Overdamped or underdamped? What bandwidth of filtering would be needed to be sure that fluid displacements of ~ 1 Å are detectable above the thermal noise of the bundle? Is this roughly consistent with the observed threshold power?

In mammalian ears, the hair cells sit on top of a structure called the basilar membrane, the tips of the stereocilia are in contact with another structure, the tectorial membrane, and the entire organ, called the cochlea, is wrapped into a spiral and embedded in bone [need a figure here!]. Sound waves impinging on the eardrum are coupled into the cochlea to produce a pressure difference across the basilar membrane, which then vibrates, ultimately causing motions of the stereocilia. Because it is surrounded by fluid, motions of neighboring pieces of the basilar membrane are coupled, and the result is a wave that travels along the membrane; because of gradations in the mechanical properties of the system, high frequency waves have their peak amplitude near the entrance to the cochlea and low frequency waves have their peak near the end or apex of the cochlea. Helmholtz knew about the structure of the inner ear, and since he saw fibrous components in the various membranes, he imagined that these might be taught, resonant strings. Because the strings were of different lengths and thicknesses, varying smoothly along the length of the cochlea, the resonant frequency would also vary. Thus, Helmholtz had the basic picture of the cochlea as a mechanical system which analyzes incoming sounds into component frequencies, sorting them to different locations along the basilar membrane. It is not clear how seriously he took the details of the mechanics, but the picture of the ear as frequency analyzer or bank of filters was taken very seriously, and indeed this picture accounts for many perceptual phenomena. The first direct measurements of basilar membrane motion were made by von Békésy, who opened the cochleae of various animals, sprinkled reflecting flakes onto the membrane, and observed its motion stroboscopically under the microscope.[57] Békésy saw the traveling wave of vibrations along the basilar membrane, and he saw the mechanical sorting of frequencies which Helmholtz had predicted.

Problem 88: Cochlear mechanics. Generate a problem that gives the students a tour of classical ideas about the traveling wave along the basilar membrane. Get them to use WKB methods to solve, understand how the peak forms etc..

Békésy was also immediately impressed with the scale of motions in the inner ear. To make the basilar membrane vibrate by $\sim 1\,\mu$m and hence be easily visible under the light microscope, he had to deliver sounds at what would be the threshold of pain, ~ 120 dB SPL.[58] If we just extrapolate linearly, $1\,\mu$m at 120 dB SPL corresponds to 10^{-12} m at 0 dB SPL, or ~ 0.01 Å (!). This is an astonishingly small displacement.

---

[57] Many of von Békésy's key contributions are collected in a volume published relatively late in his life, along with various reminiscences and quasi–philosophical remarks. As an example, he notes that in science good enemies are much more valuable than good friends, since enemies will take the time to find all your mistakes. Unfortunately, in the process of this dialogue, some of the enemies become friends and hence, by von Békésy's criteria, their usefulness is lost.

[58] SPL stands for sound pressure level. It is conventional in acoustics to measure the intensity of sounds logarithmically relative to some standard. 10 dB corresponds to a power ratio of $10\times$, so 20 dB corresponds to a factor of $10\times$ higher sound pressure variations. For human hearing the standard reference (0 dB SPL) is a pressure of 2×10^{-5} N/m^2 which is close to the threshold of hearing at frequencies near 2 kHz.

Problem 89: Brownian motion of the basilar membrane. Generate a problem that takes the students through the analysis of Brownian motion in a continuous system, with basilar membrane as an example.

Békésy also observed that the frequency selectivity of the basilar membrane motion was quite modest. More precisely, the peak of the vibrations in response to a single frequency was quite broad, spreading over a distance along the cochlea that corresponds to more than ten times the apparent bandwidth over which we integrate. This discrepancy seems to have caused more concern than the extrapolated displacement. On the one hand, if it is correct it suggests that there are mechanisms to sharpen frequency selectivity that come after the mechanics of the inner ear, perhaps at the level of neural circuitry. Békésy was very much taken with the ideas of lateral inhibition in the retina, and suggested that this might be a much more general concept for neural signal processing. On the other hand, von Békésy studied dissected cochleae that were, not to put too fine a point on it, dead. By the 1970s, it became clear that individual neurons emerging from the cochlea had frequency selectivity which was sharper than suggested by von Békésy's measurements, and that (especially in mammals) this selectivity was extremely fragile, dependent on the health of the cochlea—so much so that the tuning properties of individual neurons could be changed within minutes by blocking blood flow to the ear, recovering just as quickly when the block was relieved.

Observations on the fragility of cochlear tuning emphasized the challenge of making direct mechanical measurements on more intact preparations, and presumably at more comfortable sound levels. To make measurements of smaller displacements, a number of tools from experimental physics were brought to bear: the Mössbauer effect, laser interferometry, and Doppler velocimetry. At the same time, several groups turned to non–mammalian systems which seemed like they would be more robust, such as the frog sacculus and the turtle cochlea, and especially in these systems it proved possible to make much more quantitative measurements on the electrical responses of the hair cells and eventually on their mechanical properties. In the midst of all this progress came the most astonishing evidence for active mechanical filtering in the inner ear.

If we build an active filter via feedback, and try to narrow the bandwidth as much as possible, we are pushing the system to the edge of instability. It is not difficult to imagine that, with active feedback provided by biological mechanisms, that some sort of pathology could result in an error that pushed past the gain past the bifurcation, turning a narrow bandwidth filter into an oscillator. If incoming sounds are efficiently coupled to motions of the active elements in the inner ear, then spontaneous oscillations of these elements will couple back, and the ear will emit sound. Strange as it may seem, careful surveys show that almost half of all ears have a "spontaneous oto–acoustic emission;" a rather quiet, narrow band sound that can be detected by placing a microphone in the ear canal, as shown in Fig 78. Importantly, the statistics of the sounds being emitted are not those of filtered noise, but rather those expected from a true oscillator—the distribution of instantaneous sound pressures has a minimum at zero, as expected if the quiet state is unstable.

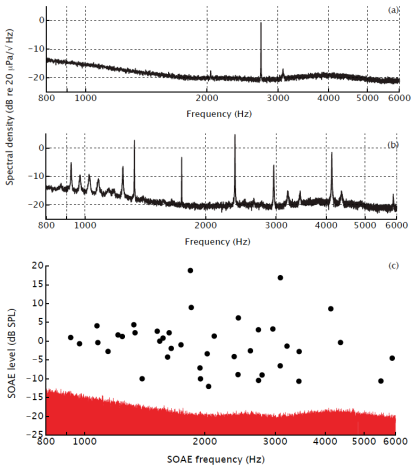

FIG. 78 Spontaneous emission of sounds from the human ear, from van Dijk et al (2011). Top panels show the spectral density of sounds in the ear canals of two subjects. Bottom panel shows the intensities and frequencies of 41 spectral peaks found in 8 subjects, compared with the noise background.

[Need to wrap this up .. Direct measurement on ciliary mechanics in different systems; violation of FDT as evidence of activity. Note re electrical resonances. Look at Marcelo & Jim's papers to see smoking gun for Hopf bifurcation.]

[Reach a conclusion!]

Now is a good time to look back at Schrödinger's remarkable little book (Schrödinger 1944). The idea which him were presented by Timoféef–Ressovsky et al (1935). For some later perspectives see Delbrück's Nobel lecture (1970); the title refers to an earlier lecture, also very much worth reading for its eloquence and prescience (Delbrück 1949). A review of DNA structure is given in Appendix A.5, and some general references on molecular biology are at the end of Section II.B. The ideas of kinetic proofreading—and, as emphasized in the text, the idea that there is a general physics problem cutting across a wide range of biological phenomena—were presented in Hopfield (1974) and Ninio (1975). Hopfield (1980) constructed a scenario in which the basic idea of paying (energetically) for increased accuracy still operates, but with none of the experimental signatures of the original proofreading scheme. [Need refs that proofreading is correct!]

Delbrück 1949: A physicist looks at biology. M Delbrück, *Trans Conn Acad Arts Sci* **38,** 173–190 (1949). Reprinted in *Phage and the Origins of Molecular Biology*, J Cairns, GS Stent & JD Watson, eds, pp 9–22 (Cold Spring Harbor Press, Cold Spring Harbor NY, 1966).

Delbück 1970: A physicist's renewed look at biology: twenty years later. M Delbrück, *Science* **168,** 1312–1315 (1970). Also available at http://nobelprize.org.

Hopfield 1974: Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. JJ Hopfield, *Proc Nat'l Acad Sci (USA)* **71,** 4135–4139 (1974).

Hopfield 1980: The energy relay: a proofreading scheme based on dynamic cooperativity and lacking all characteristic symptoms of kinetic proofreading in DNA replication and protein synthesis. JJ Hopfield, *Proc Nat'l Acad Sci (USA)* **77,** 5348–5252 (1980).

Ninio 1975: Kinetic amplification of enzyme discrimination. J Ninio *Biochimie* **57,** 587–595 (1975).

Schrödinger 1944: *What is Life?* E Schrödinger (Cambridge University Press, Cambridge, 1944).

Timoféef–Ressovsky et al 1935: Über die Natur der Genmutation und der Genstruktur. NW Timoféef–Ressovsky, KG Zimmer & M Delbrück, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen* **1,** 190–245 (1935). [Is there a translation available?]

Need refs for mutator strains, selection on error rates, tradeoff among energy costs, accuracy and speed of growth, ... The fact that some mutations lead to changes in mutation rate could have dramatic consequences for the pace of evolutionary change, as emphasized by Magnasco & Thaler (1996). [where they "right," even in outline? what has happened since?]

Magnasco & Thaler 1996: Changing the pace of evolution. MO Magnasco & DS Thaler, *Phys Lett A* **221,** 287–292 (1996).

The basic idea of kinetic proofreading—an enzymatic mechanism dissipating energy to stabilize a "better than Boltzmann" distribution of molecular states—has by now been applied in several different contexts. For the disentangling of DNA strands, see (Yan et al 1999, 2001), who were inspired in part by the experiments of Rybenkov et al (1997). For the sensitivity and specificity of initial events in the immune response, see McKeithan (1995) and Altan–Bonnet & Germain (2005), and for a more general view of signal transduction specificity see Swain & Siggia (2002).

Altan–Bonnet & German 2005: Modeling T cell antigen discrimination based on feedback control of digital ERK responses. G Altan–Bonnet & RN Germain, *PLoS Biology* **3,** 1925–1938 (2005).

McKeithan 1995: Kinetic proofreading in T–cell receptor signal transduction. TW McKeithan, *Proc Nat'l Acad Sci (USA)* **92,** 5042–5046 (1995).

Rybenkov et al 1997: Simplification of DNA topology below equilibrium values by type II topoisomerases. VV Rybenkov, C Ullsperger, AV Vologodskii & NR Cozzarelli, *Science* **277,** 690–693 (1997).

Swain & Siggia 2002: The role of proofreading in signal transduction specificity. PS Swain & ED Siggia, *Biophys J* **82,** 2928–2933 (2002).

Yan et al 1999: A kinetic proofreading mechanism for disentanglement of DNA by topoisomerases. J Yan, MO Magnasco & JF Marko, *Nature* **401,** 932–935 (1999).

Yan et al 2001: Kinetic proofreading can explain the suppression of supercoiling of circular DNAs by type–II topoisomerases. J Yan, MO Magnasco & JF Marko, *Phys Rev E* **63,** 031909 (2001).

The use of optical forces to manipulate biological systems goes back to work by Ashkin & Dziedzic (1987), which in turn grew out of Ashkin's earlier work (Ashkin 1978, 1980). It is worth noting that the same ideas of optical trapping for neutral, dielectric particles were a key step in the development of atomic cooling, as described, for example, by Chu (2002). For the state of the art in single molecule experiments, see Greenleaf et al (2007). The experiments on single RNA polymerase molecules were by Shaevitz et al (2003) and Abbondanzieri et al (2005). [Add refs for single ribosome experiments, depends on what happens in the text.]

Abbondanzieri et al 2005: Direct observation of base-pair stepping by RNA polymerase. EA Abbondanzieri, WJ Greenleaf, JW Shaevitz, R Landick & SM Block, *Nature* **438,** 460–465 (2005).

Ashkin 1978: Trapping of atoms by resonance radiation pressure. A Ashkin, *Phys Rev Lett* **40,** 729–732 (1978).

Ashkin 1980: Applications of laser radiation pressure. A Ashkin, *Science* **210,** 1081–1088 (1980).

Ashkin & Dziedzic 1987: Optical trapping and manipulation of viruses and bacteria. A Ashkin & JM Dziedzic, *Science* **235,**1517–1520 (1987).

Chu 2002: The manipulation of neutral particles. S Chu, in *Nobel Lectures, Physics 1996–2000* G Ekspong, ed (World Scientific, Singapore, 2002). Also available at http://nobelprize.org.

Greenleaf et al 2007: High–resolution, single–molecule measurements of biomolecular motion. WJ Greenleaf, MT Woodside & SM Block. *Annu Rev Biophys Biomol Struct* **36,** 171–190 (2007).

Shaevitz et al 2003: Backtracking by single RNA polymerase molecules observed at near–base–pair resolution. JW Shaevitz, EA Abbondanzieri, R Landick & SM Block, *Nature* **426,** 684–687 (2003).

The images of hair cells are from [be careful to revise with new figures] and Dallos (1984). Early experiments on the mechanics of the stereocilia are by Flock & Strelioff (1984; Strelioff & Flock 1984). The remarkable vibration sensitivity of the frog inner ear is described in two papers, Narins & Lewis (1984) and Lewis & Narins (1985). Estimates of the power flowing into the inner ear at threshold, as well as the minimum detectable power in other sensory systems, were given some time ago by Khanna & Sherrick (1981). [I think there is also a nice reference to threshold powers in relation to electroreceptors; try to find this.]

Dallos 1984: Peripheral mechanisms of hearing. P Dallos, in *Comprehensive Physiology 2011, Supplement 3: Handbook of Physiology, The Nervous System, Sensory Processes* pp. 595-637 (Wiley–Blackwell, 1984).

Flock & Strelioff 1984: Studies on hair cells in isolated coils from the guinea pig cochlea. Å Flock & D Strelioff, *Hearing Res* **15,** 11–18 (1984).

Khanna & Sherrick 1981: The comparative sensitivity of selected receptor systems. SM Khanna & CE Sherrick, in *The Vestibular System: Function and Morphology,* T Gualtierotti, ed, pp 337–348 (Springer–Verlag, New York, 1981).

Lewis & Narins 1985: Do frogs communicate with seismic signals? ER Lewis & PM Narins, *Science* **227,** 187–189 (1985).

Narins & Lewis 1984: The vertebrate ear as an exquisite seismic sensor. PM Narins & ER Lewis, *J Acoust Soc Am* **76,** 1384–1387 (1984).

Strelioff & Flock 1984: Stiffness of sensory–cell hair bundles in the isolated guinea pig cochlea. D Strelioff & Å Flock, *Hearing Res* **15,** 19–28 (1984).

[Find early references to active cooling ...] Recent examples of active cooling are by Corbitt et al (2007) and Abbott et al (2009), who are aiming at improving the sensitivity of gravitational wave detection. For discussion of the quantum limits to mechanical measurements see Caves et al (1980), Caves (1985), and Braginsky & Khalili (1992). For the quantum limits to amplifier noise (which has a long history), see Caves (1982). For discussions of thermal noise, I have always liked the treatment in Kittel's little book, cited at the end of Section 2.1, and this includes a discussion of the "thermal noise power."

Abbott et al 2009: Observation of a kilogram–scale oscillator near its quantum ground state. B Abbott et al (LIGO collaboration), *New J Phys* **11,** 073032 (2009).

Braginksy & Khalili 1992: *Quantum Measurement.* VB Braginsky & F Ya Khalili (Cambridge University Press, Cambridge, 1992).

Caves 1982: Quantum limits on noise in linear amplifiers. CM Caves, *Phys Rev D* **26,** 1817–1839 (1982).

Caves 1985: Defense of the standard quantum limit for free-mass position. CM Caves, *Phys Rev Lett* **54,** 2465–2468 (1985).

Caves et al 1980: On the measurement of a weak classical force coupled to a quantum–mechanical oscillator. I: Issues of principle. CM Caves, KS. Thorne, RWP Drever, VD Sandberg & M Zimmermann, *Revs Mod Phys* **52,** 341–392 (1980).

Corbitt et al 2007: Optical dilution and feedback cooling of a gram–scale oscillator to 6.9 mK. T Corbitt, C Wipf, T Bodiya, D Ottaway, D Sigg, N Smith, S Whitcomb & N Mavalvala, *Phys Rev Lett* **99,** 160801 (2007).

The physics of hearing is a subject that goes back to Helmholtz (1863) and Rayleigh (1877). An early measurement using tone–on–tone masking to probe the sharpness of frequency selectivity in the cochlea was done by Wegel & Lane (1924); a modern account of psychophysical measurements on effective detection bandwidths is given by [find ref!]. The classical measurements on cochlear mechanics by von Békésy were collected in 1960. The modern era of such measurements begins with the use of the Mössbuaer effect (Frauenfelder 1962) to measure much smaller displacements of the basilar membrane (Johnstone et al 1967), and to demonstrate these motions have decided sharper frequency tuning in response to quieter sounds (Rhode 1971). These experiments triggered renewed interest in theories of cochlear mechanics, and some beautiful papers from this period are by Zweig et al (1976; Zweig 1976) and by Lighthill (1981), both of which connect the dynamics of the inner ear to more general physical considerations. There then followed a second wave of mechanical measurements, using interferometry (Khanna & Leonard 1982), Doppler velocimetry [Ruggero?], and improved Mössbauer methods (Sellick et al 1982). Some perspective on these measurements and models as they emerged can be found in Lewis et al (1985), which also makes connections to the mechanics of other inner ear organs. [Something more modern?]

Békésy 1960: *Experiments in Hearing* G von Békésy, EG Wever, ed (McGraw–Hill, New York, 1960).

Frauenfelder 1962: *The Mössbauer Effect* (WA Benjamin, New York, 1962).

Helmholtz 1863: *Die Lehre den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* H von Helmholtz (Vieweg und Sohn, Braunschweig, 1863). The most widely used translation is the second English edition, based on the fourth (and last) German edition of 1977; translated by AJ Ellis with an introduction by H Margenau, *On the Sensations of Tome as a Physiological Basis for the Theory of Music* (Dover, New York, 1954).

Johnstone et al 1967: Basilar membrane vibration examined with the Mössbauer technique. BM Johnstone & AJF Boyle, *Science* **158,** 390–391 (1967).

Khanna & Leonard 1982: Basilar membrane tuning in the cat cochlea. SM Khanna & DGB Leonard, *Science* **215,** 305–306 (1982).

Lewis et al 1985: *The Vertebrate Inner Ear* ER Lewis, EL Leverenz & WS Bialek (CRC Press, Boca Raton FL, 1985).

Lighthill 1981: Energy flow in the cochlea. J Lighthill, *J Fluid Mech* **106,** 149–213 (1981).

Rayleigh 1877: *Theory of Sound* JW Strutt, Baron Rayleigh (Macmillan, London, 1877). More commonly available is the revised and enlarged second edition, with a historical introduction by RB Lindsay (Dover, New York, 1945).

Rhode 1971: Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. WS Rhode, *J Acoust Soc Am* **49,** 1218– 1231 (1971).

Sellick et al 1980: Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique. PM Sellick, R Patuzzi & BM Johnstone, *J Acoust Soc Am* **72,** 131–141 (1982).

Wegel & Lane 1924: The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. RL Wegel & CE Lane, *Phys Rev* **23,** 266–285 (1924).[59]

Zweig 1976: Basilar membrane motion. G Zweig, *Cold Spring Harb Symp Quant Biol* **40,** 619–633 (1976).

Zweig et al 1976: The cochlear compromise. G Zweig, R Lipes & JR Pierce, *J Acoust Soc Am* **59,** 975–982 (1976).

The idea of active filtering in the inner ear goes back to a remarkably prescient paper by Gold (1948), who is better known, perhaps, for his contributions to astronomy and astrophysics; see Burbidge & Burbidge (2006). The idea that active elements are at work in the mechanics of the mammalian cochlea gained currency as experiments showed the "vulnerability" of frequency selectivity (Evans 1972), and with the dramatic observation of acoustic emissions from the ear (Kemp 1978, Zurek 1981); the data shown in Fig 78 are from van Dijk et al (2011). Importantly these emissions are observed not only from the rather complex mammalian cochlea, but from simpler ears of amphibians [get proper refs from van Dijk et al]. [Need a recent overview of acoustic emissions.] The idea that active filtering is essential for noise reduction is discussed in Bialek (1987). The modern view of active filtering as an approach to the Hopf bifurcation begins with Eguíliz et al (2000), and has been developed by [...].

Bialek 1987: Physical limits to sensation and perception. W Bialek, *Annu Rev Biophys Biophys Chem* **16,** 455–478 (1987).

Burbidge & Burbidge 2006: Thomas Gold, 1920–2004. G Burbidge & M Burbidge, *Bio Mem Nat'l Acad Sci (USA)* **88,** 1–15 (2006).

van Dijk et al 2011: The effect of static ear canal pressure on human spontaneous otoacoustic emissions: Spectral width as a measure of the intra–cochlear oscillation amplitude. P van Dijk, B Maat & E de Kleine, *J Assoc Res Otolaryngol* **12,** 13–28 (2001).

Eguíliz et al 2000: Essential nonlinearities in hearing. VM Eguíliz, M Ospeck, Y Choe, AJ Hudspeth & MO Magnasco, *Phys Rev Lett* **84,** 5232–5235 (2000).

---

[59] It is amusing to note that this paper is sometimes cited in the biological literature as having been published in the journal *Physiological Reviews*. Presumably this reflects authors or editors copying the reference to *Phys Rev* and "correcting" it to *Physiol Rev* without checking.

Evans 1972: The frequency response and other properties of single fibres in the guinea–pig cochlear nerve. EF Evans, *J Physiol (Lond)* **226,** 263–287 (1972).

Gold 1948: Hearing. II: The physical basis of the action of the cochlea. T. Gold, *Proc R Soc Lond Ser B* **135,** 492–498 (1948).

Kemp 1978: Stimulated acoustic emissions from within the human auditory system. DT Kemp, *J Acoust Soc Am* **64,** 1386–1391 (1978).

Zurek 1981: Spontaneous narrowband acoustic signals emitted by human ears. PM Zurek, *J Acoust Soc Am* **69,** 514–523 (1981).

E. Perspectives

Many of life's phenomena exhibit a startling degree of reliability and precision. Organisms reproduce and develop with surprising predictability, and our own perceptual experience of the world feels certain and solid. On the other hand, when we look inside a single cell, or even at the activity of single neurons in brain, things look very noisy. Are the building blocks of biological behavior really so noisy? If so, how can we understand the emergence of reliability and certainty from all this mess?

Many of the problems faced by living organisms can be phrased as sensing, processing and responding to signals. If we look at a part of a system involved in such sensory tasks, we have to be careful in assessing noise levels. As a simple example, if we build a system in the lab that measures a small signal, and somewhere in this system there is an amplifier with very high gain, then surely we will find places in the circuitry where the voltage fluctuations are very large. Alternatively, there might be no gain, just a lot of noise. Thus, the variance of the noise at one point in the system, by itself, tells us nothing about its true degree of noisiness.

When we build sensors in the lab, we measure their noise performance by referring the noise to the input—estimating the noise level that would have to be added to the signals that we are trying to sense so as to account for the noise that we see at the output. This effective noise level is also the noise that limits the detectability of small signals, or the discriminability of signals that are very similar to one another. Importantly, for many sensors there are physical limits on this effective noise at the input, which allows us to put the noise performance on an absolute scale.

What we have done in this Chapter is to look at several instances it which it has been possible to carry out the program of "referring noise to the input" for increasingly complex biological systems. This is by far not a closed subject, and it is a minority of systems that have been analyzed in this way. Nonetheless, it is striking that, in so many disparate instances, the noise performance of biological systems indeed is close to the relevant physical limits. This of course is in the spirit of what we learned from the case of photon counting in vision, but it seems much more general.

[I need to give some exegesis of this, and what it implies. Perhaps because I have spent so much time on these issues myself, I am having difficulty at the moment generating enough distance to be clear and objective (and not just to repeat what was said at the end of the previous chapter). So, I will need to come back to this. Sorry to leave things hanging in an important spot!]