Kinetic proofreading and sensitivity of steady states

Jeremy A. Owen
(Dated: May 18, 2018)

Many fundamental processes in biology, like DNA replication and protein synthesis, are surprisingly precise despite thermal noise being significant at the relevant scales. Part of this is due to nonequilibrium processes, like kinetic proofreading, that can amplify a small difference in binding affinity. Recent work has studied multi-step generalizations of proofreading through the lens of a quantity called "discriminatory index". We note this is closely related to the sensitivity of the steady state density of a Markov chain to perturbations of rates, a perspective that emphasizes the generality of the underlying ideas, making especially clear the close analogy between kinetic proofreading and the nonequilibrium sensitivity of biological sensors.

I. INTRODUCTION

Many biological recognition processes are far more accurate than permitted at equilibrium. A classic example is translation, where the difference between the binding of the right or wrong tRNA to a codon can be a single hydrogen bond, and a typical error rate of the process is 10^{-4} errors per codon. At equilibrium, achieving this specificity would require a free energy difference of $\sim 9k_{\mathrm{B}}T$ which is considerably in excess of what is typical for a hydrogen bond.

It is easy to imagine a nonequilibrium scheme to improve on this error rate—one could use a chemical potential difference to selectively drive the "right" reaction. But on reflection this is completely unsatisfactory, as it merely shifts the challenge of discrimination to whatever is selectively coupling e.g. ATP hydrolysis to a reaction involving only the "right" tRNA.

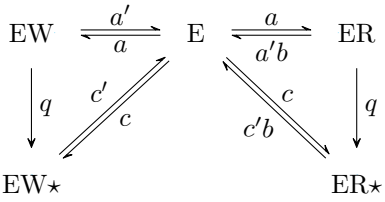
What we need instead is a process that *amplifies* a small difference in binding affinity, while treating both substrates completely symmetrically. This is what kinetic proofreading [1, 2] does.

The organization of this paper is as follows. First, we describe the original kinetic proofreading scheme, and generalizations involving multiple proofreading steps. Then we set out some basic facts about Markov chains, and explain how they constrain the discriminatory ability of general proofreading schemes. Finally, in the discussion, we will describe the close link between kinetic proofreading and nonequilibrium sensing—made especially clear by the perspective we take here—and touch on what is known from experiments about error correction in real biological systems.

A. Hopfield's kinetic proofreading

Hopfield [1] (and independently Ninio [2]) proposed the following scheme that can improve an error rate b to b^2. Consider an enzyme E with two very similar substrates—a "right" one R and a "wrong" one W. Suppose the concentrations of both are similar and both are very large compared to that of E. At equilibrium (e.g. in the absence of any ATP consumption), one can show that the relative probability of finding the enzyme bound to one substrate rather than the other is $b = \frac{[EW]}{[ER]} = e^{-\Delta}$. where $\Delta > 0$ is the free energy difference between the two enzyme-substrate complexes EW and ER (in units of $k_{\mathrm{B}}T$).

Consider the following chemical reaction network, the "butterfly graph" [3], in which two transitions are driven close to irreversibility at rate q. Note the symmetry of this network, broken only by the appearance of the equilibrium error fraction b on the right hand side.



At steady state, we have

$$\eta \equiv \frac{[EW\star]}{[ER\star]} = \frac{b(a'b+q)(a'c+(a+c)q)}{(a'+q)(a'bc+(a+c)q)}$$

which tends to b^2, a significant improvement over the equilibrium error rate, if we send c to zero and *then* q to zero. But note that the order in which limits are taken here matters—if we take them in the other order we get b—no improvement over equilibrium. This reveals subtlety in this limiting procedure, which is discussed further in [3].

There is however a simple intuition underlying this scheme. If we imagine only $ER\star$ and $EW\star$ can release product, and that the on-rate c is small, the substrate R or W has two "chances" to dissociate before the enzyme is "finished" acting, at rates suppressed by a factor of b in the case of R [1, 4].

B. Generalizations

Can one do better than b^2 with the network above in a different asymptotic limit of rates? Recent work [5] introduced the *discriminatory index* ν (there is a sign difference due to choice of convention):

$$\nu = \frac{\partial \log(\eta)}{\partial \Delta}$$

A scheme that is able to improve error fraction to b^m must have $\nu = m$. Hopfield's scheme in the limit we described above has $\nu = 2$. In general ν will depend on Δ and what limit of the other parameters we are considering [5], but cannot exceed 2 for this network.

To achieve better discrimination, we can introduce further activated intermediates (e.g. $ER\star\star$) that can also dissociate to E [5]. This introduces further more transition rates that depend on Δ. We will see this is key, and that in general networks ν is bounded in terms of the number of edges involving Δ.

II. FACTS ABOUT MARKOV CHAINS

One interpretation of the butterfly graph is that describes possible transitions of a single enzyme, occurring stochastically with rates specified as edge labels. In other words, as a continuous-time Markov chain—a stochastic processes with finite state space whose future evolution depends on the past only through the present. In a more general case, at time t, the system is found in the ith of its N states with probability $p_i(t)$, and

$$\dot{p} = Mp \tag{1}$$

where M is some matrix whose column sums are zero at all times, so that probability is conserved. Off-diagonal entries represent transition rates so they must be non-negative[1].

The error rate η in proofreading is the ratio of two components of p after a long time. In the rest of this section we give some simple corollaries of well-known results about Markov chains that will be useful in our discussion.

A. Steady state density

Consider the graph G associated to a Markov chain with rate matrix M that has a vertex for each state and an edge for each possible transition. If there is a path in G from any state to any other, the master equation (1) has a unique steady state $p_{\rm ss}$, given explicitly by the matrix tree theorem (MTT) [7–9] in terms of the directed spanning trees of G

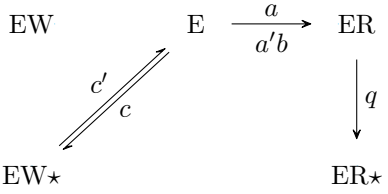
$$p_{{\rm ss},x} \propto \sum_{\substack{\text{spanning trees}\\ \text{directed to x}}} \prod_{\text{tree edges } i \to j} M_{ji}$$

A useful corollary is given by [10]—the quantity $\log\left(p_{{\rm ss},i}/p_{{\rm ss},j}\right)$ is bounded above and below by the maximum and minimum, over all (non-self-intersecting) paths $j \to 1 \to \cdots \to m \to i$ in G of the quantity

$$\Delta S_{\rm env} = \log\left(\frac{M_{1j}\dots M_{im}}{M_{mi}\dots M_{j1}}\right)$$

which can be identified under many circumstances with the change in entropy of the environment when the system follows the path $j \to \cdots \to i$ [10, 11]. At equilibrium, this quantity equals $\beta\Delta E$ along *any* path between two states of energy difference ΔE, but out of equilibrium it can be path-dependent.

The $\log\left(p_{{\rm ss},i}/p_{{\rm ss},j}\right)$ is exactly the sort of quantity we wish to extremize in kinetic proofreading, and one might imagine that proofreading works by supplying a path between $EW\star$ and $ER\star$ with a very large value of $\Delta S_{\rm env}$, such as:



But, just like the idea mentioned in the introduction of selectively driving $E \to ER$, making this path the dominant one requires an asymmetry that cannot be achieved without first solving the discrimination issue.

B. Response of the density to perturbations

If we change some transition rates M_{ij}, how does the steady state distribution $p_{\rm ss}$ respond?

The most intuitive thing we can say is that if a rate $i \to j$ is increased, the steady state probability of being in state i must decrease, and that of being in j should increase.[2] Less obvious perhaps is the fact that the steady state probability of any state is *monotonic* in each rate constant.[3]

These results hold for arbitrarily large changes of rate constants. What will be most relevant for what follows, however, is the following bound on the logarithmic sensitivity of the steady state density to changes in transition rates,

$$\left|\frac{\partial \log(p_{{\rm ss},k})}{\partial \log(M_{ij})}\right| \leq 1. \tag{2}$$

[1] Alternatively, we can view (1) as the rate equation of a monomolecular chemical reaction network with mass action kinetics [6], where p is to be interpreted as a vector of concentrations. Although it has a very different origin, arising in a large N limit from a different stochastic process, the equation is the same.

[2] The former result follows immediately from MTT since rates of transitions from i do not appear in the numerator of p_i. The argument for the latter seems more involved, an asymmetry I would like to understand better.

[3] By MTT we have (for some positive a, b, c, and d independent of k):

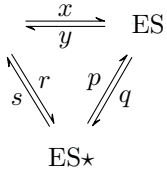
$$p_{{\rm ss},i} = \frac{a+bk}{c+dk}, \qquad \frac{dp_{\rm ss}}{dk} = \frac{bc-ad}{(c+dk)^2}.$$

This result is a consequence of the multilinearity (linear in each rate constant separately) of the numerator and denominator of the expressions given by MTT for $p_{\mathrm{ss},k}$ (see Appendix for proof).

Eq. (2) is stated for discrete time Markov chains and proved using a different approach in [12]. I believe that version implies the continuous time result given here.

III. DISCRIMINATION AND SENSITIVITY

Consider the network:



In steady state, how sensitive is the ratio $[E]/[ES\star]$ to changes in the rates, for example, s and y? At first, this does *not* seem so related to the challenge addressed by kinetic proofreading—how to minimize error in a discrimination task—after all where even is the "right" and "wrong" substrate in this description? But in fact we will see these questions are extremely close.

First, we note that the error fraction from the example in the introduction can be written [3, 5]:

$$\eta = \frac{[EW\star]}{[E]}\frac{[E]}{[ER\star]}$$

The binding energy difference affects only the second of these factors (which one is matter of convention). This implies that the discriminatory index is

$$\nu = \frac{\partial \log(\eta)}{\partial \Delta} = \frac{\partial \log([E]/[ER\star])}{\partial \Delta}$$

Furthermore, the ratio $[E]/[ER\star]$ only depends on quantities to the "right" of E in the "butterfly graph" [3]. So it is sufficient to consider the triangle network given at the beginning of this section when computing ν.

A. A weak bound relevant to proofreading

Now imagine we have a completely general network (not necessarily the triangle above) and suppose the binding energy difference Δ appears in n different rates, in the exponent. The quantity ν of a proofreading scheme based on such a network is then of the form (for some i and j)

$$\nu = \frac{\partial \log(p_{\mathrm{ss},i}/p_{\mathrm{ss},j})}{\partial \Delta} = \sum_{k=1}^{n} \frac{\partial M_{m(k)n(k)}}{\partial \Delta}\frac{\partial \log(p_{\mathrm{ss},i}/p_{\mathrm{ss},j})}{\partial M_{m(k)n(k)}}$$

which implies

$$|\nu| \leq \sum_{k=1}^{n} \left|\frac{\partial \log(p_{\mathrm{ss},i}/p_{\mathrm{ss},j})}{\partial \log(M_{m(k)n(k)})}\right| \leq 2n \qquad (3)$$

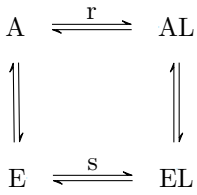
where the triangle inequality has been used twice (once for the sum and on $|\log(p_{\mathrm{ss},i}/p_{\mathrm{ss},j})| = |\log(p_{\mathrm{ss},i}) - \log(p_{\mathrm{ss},j})|$), and the bound Eq. (2) has been used once. This inequality is not tight—Hopfield's network has $n = 2$ but $\nu \leq 2$. Tighter bounds are given in [5]. But indeed we have not used all the usual constraints to get this result, for example that the binding energy Δ only appear in *dissociation* rates, as is commonly supposed. We also note that the inequality (3) above does not preclude $\nu < 1$ (or indeed for ν to be negative), and indeed such "anti-proofreading" [5] is possible.

IV. DISCUSSION

A. Relation to nonequilibrium sensing

There is a close analogy between kinetic proofreading and the "sensing" capabilities of biological circuits, for example the chemotaxis system in *E. coli* [13]. In both cases, one desires that a small change in a transition rate (due to a change in binding affinity or a different attractant concentration outside the cell, respectively) to have a large effect, and one can do better out of equilibrium.

The authors of [13] study a four state model for a receptor that can be either active (A) or inactive (I) and either have a ligand L bound or not.



Imagine that when the receptor is active, it promotes a useful response to the ligand (tumbling in *E. coli*). A figure of merit is then

$$R' = \frac{\partial \log((p_{\mathrm{A}} + p_{\mathrm{AL}})/(p_{\mathrm{E}} + p_{\mathrm{EL}}))}{\partial \log([L])}$$

(which bounds above the sensitivity called R in [13]). The ligand concentration $[L]$ appears (multiplicatively) in the rates r and s, so R' is a quantity very much like ν for kinetic proofreading. The same arguments we gave earlier (this time also using the convexity of log) give a bound $|R'| \leq 8$ (there are two rates containing $[L]$, and a ratio of two sums of two probabilities). Once again, we note this is a very loose bound.

B. Experimental evidence for proofreading

Early evidence for kinetic proofreading came from a process upstream from translation—the charging of tRNAs with amino acids. The observation was that much more ATP was consumed when the incorrect amino acid was supplied compared to the correct one [14]. This is

consistent with proofreading because the wrong substrate is likely to fall off many times before it is incorporated, and each "attempt" involves ATP hydrolysis.

More recently, it has been possible using FRET, which can be used to detect specific bimolecular interactions, to uncover mechanistic details of translation providing strong support for kinetic proofreading [15]. The same technology has been used to quantify the role of dissipation (measuring the formation of a complex involving a kinase) in the adaptation accuracy of chemotaxis in *E. coli* (how precisely the tumble rate returns to its original value after exposure to attractant) [16]. Much like the sensitivity we discussed above (a different figure of merit for the same sensing system), high adaptation accuracy depends on the system being out of equilibrium.

Another consideration relevant in real systems is *speed*. If the enzyme is to complete its functional role, for example a polymerase that must move on to the next nucleotide, it must eventually release product. We can model this with irreversible "catalytic" transitions from $ER\star$ and $EW\star$. Only in the limit where these rates are very small compared to other rates can the proofreading limits describe here be approached. This means accuracy imposes a time cost. There is evidence [17] that in fact real error correcting systems are tuned to maximize speed instead of accuracy (once the error is below a certain threshold).

V. CONCLUSION

A general bound on the sensitivity of the steady states of Markov chains is closely related to limits on discrimination imposed by structural features of reaction networks. This bound applies to models of kinetic proofreading, proofreading with multiple steps, and nonequilibrium sensing, highlighting the underlying similarity of these processes which are important to the processes of the central dogma and in other biological contexts, like in bacterial chemotaxis.

[1] J. J. Hopfield, Proceedings of the National Academy of Sciences **71**, 4135 (1974).
[2] J. Ninio, Biochimie **57**, 587 (1975).
[3] F. Wong, A. Amir, and J. Gunawardena, arXiv preprint arXiv:1710.06038 (2017).
[4] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC Mathematical and Computational Biology (Taylor & Francis, 2006).
[5] A. Murugan, D. A. Huse, and S. Leibler, Physical Review X **4**, 021016 (2014).
[6] I. Mirzaev and J. Gunawardena, Bulletin of mathematical biology **75**, 2118 (2013).
[7] W. Tutte, in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 44 (Cambridge University Press, 1948) pp. 463–482.
[8] T. L. Hill, Journal of theoretical biology **10**, 442 (1966).
[9] V. Anantharam and P. Tsoucas, Statistics & Probability Letters **8**, 189 (1989).
[10] C. Maes and K. Netočnỳ, in *Annales Henri Poincaré*, Vol. 14 (Springer, 2013) pp. 1193–1202.
[11] M. Esposito and C. Van den Broeck, Physical Review E **82**, 011143 (2010).
[12] E. Thiede, B. Van Koten, and J. Weare, SIAM Journal on Matrix Analysis and Applications **36**, 917 (2015).
[13] D. Hartich, A. C. Barato, and U. Seifert, New Journal of Physics **17**, 055026 (2015).
[14] J. Hopfield, T. Yamane, V. Yue, and S. Coutts, Proceedings of the National Academy of Sciences **73**, 1164 (1976).
[15] S. C. Blanchard, R. L. Gonzalez Jr, H. D. Kim, S. Chu, and J. D. Puglisi, Nature Structural and Molecular Biology **11**, 1008 (2004).
[16] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu, Nature physics **8**, 422 (2012).
[17] K. Banerjee, A. B. Kolomeisky, and O. A. Igoshin, Proceedings of the National Academy of Sciences **114**, 5183 (2017).

* Note that I originally created the figure of the butterfly graph on page 1 for a problem in Systems Biology (8.591) when I was the TA, so it appeared on a problem set for that class.

Appendix A: Proof of Eq. (2)

Claim 1. *Let M be the rate matrix of a continuous-time Markov chain whose associated graph is strongly connected, and $p(M)$ the unique element of* $\ker M$ *whose entries sum to* 1*, viewed as a function of the entries of M. Then for any i, j $(i \neq j)$, and k:*

$$\left|\frac{\partial p_k}{\partial M_{ij}}\right| \leq \frac{p_k}{M_{ij}}.$$

Proof. We apply the matrix tree theorem. Let v_i be the sum of spanning tree monomials of M associated with state i, so e.g. $p_i = \frac{v_i}{\sum_j v_j}$. We have,

$$\left|\frac{\partial p_k}{\partial M_{ij}}\right| = \left|\frac{\frac{\partial v_k}{\partial M_{ij}}\sum_n v_n - v_k\sum_n \frac{\partial v_n}{\partial M_{ij}}}{\left(\sum_n v_n\right)^2}\right| = \frac{1}{\sum_n v_n}\left|\frac{\partial v_k}{\partial M_{ij}} - v_k\left(\sum_n \frac{\frac{\partial v_n}{\partial M_{ij}}}{\sum_m v_m}\right)\right|.$$

For any M_{ij} and n, v_n can be written as the sum of something linear in M_{ij} and something independent of M_{ij}, so we have $\frac{\partial v_n}{\partial M_{ij}} \leq \frac{v_n}{M_{ij}}$. So (using the fact that for $x, y > 0$, $|x - y| \leq \max(x, y)$),

$$\left|\frac{\partial p_k}{\partial M_{ij}}\right| \leq \frac{1}{\sum_n v_n}\max\left(\left|\frac{\partial v_k}{\partial M_{ij}}\right|, v_k\left|\sum_n \frac{\frac{\partial v_n}{\partial M_{ij}}}{\sum_m v_m}\right|\right) \leq \frac{v_k}{M_{ij}\sum_n v_n} = \frac{p_k}{M_{ij}},$$

as desired. □