Stochastic Proofreading Mechanism Alleviates Crosstalk in Transcriptional Regulation

Sarah A. Cepeda-Humerez, Georg Rieckh, and Gašper Tkačik
Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria
(Received 24 April 2015; published 8 December 2015)

Gene expression is controlled primarily by interactions between transcription factor proteins (TFs) and the regulatory DNA sequence, a process that can be captured well by thermodynamic models of regulation. These models, however, neglect regulatory crosstalk: the possibility that noncognate TFs could initiate transcription, with potentially disastrous effects for the cell. Here, we estimate the importance of crosstalk, suggest that its avoidance strongly constrains equilibrium models of TF binding, and propose an alternative nonequilibrium scheme that implements kinetic proofreading to suppress erroneous initiation. This proposal is consistent with the observed covalent modifications of the transcriptional apparatus and predicts increased noise in gene expression as a trade-off for improved specificity. Using information theory, we quantify this trade-off to find when optimal proofreading architectures are favored over their equilibrium counterparts. Such architectures exhibit significant super-Poisson noise at low expression in steady state.

DOI: 10.1103/PhysRevLett.115.248101 PACS numbers: 87.16.Yc, 87.18.Tt

In prokaryotes, transcription factors recognize and bind specific DNA sequences $L = 10$–20 base pairs (bp) in length, usually located in promoter regions upstream of the regulated genes [1]. Regulation by a single transcription factor protein (TF), or a small number of TFs interacting cooperatively, is sufficient to quantitatively account for the experimental measurements of gene expression [2], as well as to explain how any gene can be individually "addressed" and regulated only by its cognate TFs [3], without much danger of regulatory crosstalk. In eukaryotes, however, TFs seem to be much less specific ($L = 5$–10 bp, perhaps due to evolvability constraints [4], but the total genome size is larger than in prokaryotes by $\sim 10^3$) [3,5], binding promiscuously to many genomic locations [6], including to their noncognate binding sites [7]. What are the implications of this reduced specificity for the precision of gene regulation?

Thermodynamic models of regulation postulate that the rate of target gene expression is given by the equilibrium occupancy of various TFs on the regulatory sequence [8,9], and the success of this framework in prokaryotes [10] has prompted its application to eukaryotic—specifically, metazoan—enhancers [11–13]. To illustrate the crosstalk problem in this setting, consider the ratio σ of the dissociation constants to a nonspecific and a specific site for a eukaryotic TF; typically, $\sigma \sim 10^3$ (corresponding to a difference in binding energy of $\sim 7 k_B T$) [7,14]. Because there are $\nu \sim 10^2$–10^3 different TF species in a cell, TFs nonspecific to a given site will greatly outnumber the specific ones. For an isolated binding site, this would imply roughly equal occupancy by cognate and noncognate TFs, suggesting that crosstalk could be acute. For multiple sites, cooperative binding is known for its role in facilitating sharp and strong gene activation, even with cognate TFs of intermediate specificity—but could the same mechanism also alleviate crosstalk? First, there exist well-studied TFs which do not bind cooperatively (e.g., Ref. [15]). Second, to reduce crosstalk, cooperativity needs to be strong and specific, stabilizing only the binding of *cognate* TFs [16]; many proposed mechanisms lack such specificity (e.g., Refs. [17,18]). Third, even when cooperative interactions are specific, crosstalk can pose a serious constraint. Regulating a gene implies varying the cognate TF concentration throughout its dynamic range, and when this concentration is low and the target gene should be uninduced, cooperativity cannot prevent the erroneous induction by noncognate TFs. For that, the cell could keep the genes inactive by either binding specific repressors or making the whole gene unavailable for transcription. The first strategy seems widely used in bacteria, but less so in eukaryotes; the second strategy ("gene silencing") is widespread in eukaryotes, but it only happens at a slow time scale and involves a complex series of nonequilibrium steps.

Here, we propose a plausible and fast molecular mechanism which alleviates the effects of crosstalk; a detailed account of when crosstalk poses a severe constraint for gene regulation will be presented elsewhere [16]. The proposed mechanism is consistent with the known tight control over which genes are expressed in different conditions or tissues (e.g., during development [19]) on the one hand and, on the other, explains the high levels of measured noise in transcription initiation of active genes [20,21].

The simplest proofreading architecture for transcriptional gene activation that can cope with erroneous binding is presented in Figs. 1(a) and 1(b), motivated by a scheme first proposed by Hopfield [22]. Specificity is only conveyed by differential rates of TF unbinding ("off rates" k_-^{c}, k_-^{nc}, with $\sigma = k_-^{\mathrm{nc}}/k_-^{\mathrm{c}}$). There are ν noncognate TF species whose typical concentration we take to be $c_{\mathrm{nc}} = \frac{1}{2}\nu C$, and

C is the maximal concentration for the cognate TFs c_c, $c_c \in [0, C]$. The ratio $\Lambda = \nu/\sigma$ determines the severity of the crosstalk, which is weak for $\Lambda \ll 1$ and strong for $\Lambda \gg 1$. The response of the promoter to the dimensionless input concentration c [$= k_+ c_c/d$; see Fig. 1(b)] of cognate TFs is captured by the steady-state distribution of messenger RNA (mRNA), $P(m|c)$; the spread of this distribution is due to the stochasticity in gene expression, which includes random switching between promoter states and the birth-death process of mRNA expression [23]. If the reaction rates are known, $P(m|c)$ is computable from the chemical master equation corresponding to the transition

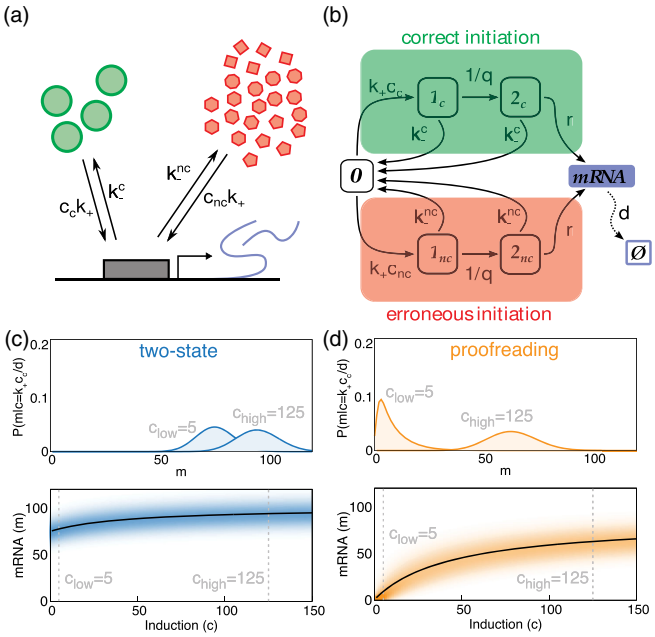

FIG. 1 (color online). (a) A schematic of cognate (the green circles) and ν kinds of noncognate (the various red shapes) TFs binding to a gene regulatory element on the DNA (the gray box), to control the mRNA expression level. (b) Transition state diagram for the proofreading gene regulation. The regulatory element can cycle between an empty state (0) and a state occupied by either a cognate (1_c) or a noncognate (1_{nc}) TF; to initiate gene expression, a further nonequilibrium transition into 2 states (with rate $1/q$) is required, driven by, e.g., hydrolysis of ATP. mRNA is expressed at rate r and degraded with rate d, the slowest process that sets our unit for time. In this figure we use $r/d = 100$, $k_-^{nc}/d = 2500$, $\sigma = 500$, $\nu = 50$, $\Lambda = \nu/\sigma = 0.1$; dimensionless concentration is $c = k_+ c_c/d$. (c),(d) Steady-state mRNA distributions for low and high concentrations of the cognate TF, c. As $qd \to 0$ (c), the proofreading model reduces to the two-state model of gene expression [24]; here, noncognate TFs initiate transcription at a high rate even when c is low, causing overlapping output distributions (blue; top panel) and a small dynamic range [the black line $= \langle m(c) \rangle$, the blue shade $= \sigma_m(c)$; bottom panel]. Proofreading (d) suppresses erroneous initiation, leading to separable output distributions (orange; top panel) and higher dynamic range (bottom panel).

diagram in Fig. 1(b); using finite-state truncation, this becomes a linear problem that is numerically tractable.

Figures 1(c) and 1(d) each compare the steady-state distributions of mRNA at low and high concentrations of cognate TF, c. The behavior crucially depends on the out-of-equilibrium rate qd. When $qd \to 0$, the scheme of Fig. 1(b) becomes a normal two-state promoter as the states 1_c and 2_c (likewise 1_{nc} and 2_{nc}) fuse into a single state. In this limit, the effect of crosstalk is highly detrimental already at $\Lambda = 0.1$ used in this example: at low c, the promoter repeatedly cycles through erroneous initiation and the gene is highly expressed at both low c and high c (where most of the expression is indeed due to correct initiation); as a result, the distributions $P(m|c)$ show substantial overlap in the two input conditions shown in Fig. 1(c). In contrast, for a nontrivial choice of q ($k_-^c \ll 1/q \simeq k_-^{nc}$), the model can exhibit proofreading. Even at low cognate concentration c, the slow, irreversible transition ensures that noncognate TFs unbind from the promoter and that erroneous initiation is consequently rare, which is manifested as a sharp peak of $P(m|c_{low})$ at small m in Fig. 1(d). The proofreading architecture generates a larger output dynamic range and consequently makes the responses distinguishable.

What are the costs to the cell of the proposed proofreading mechanism? First, the mechanism requires an energy source, e.g., adenosine triphosphate (ATP), to break detailed balance, but this metabolic burden seems negligible compared to the processive cost of transcription and translation. Second, however, is an indirect cost in terms of gene expression noise. While proofreading decreases erroneous induction, it takes longer to traverse the state transition diagram from empty state 0 to expressing state 2, and since the promoter can perform aborted erroneous initiation cycles, the fluctuations in the time to induction will also increase [25]. This will result in additional variance in the mRNA copy number at steady state compared to the two-state ($qd \to 0$) scheme. While the speed/specificity trade-off in protein synthesis has been examined before using deterministic chemical kinetics [26], this stochastic formulation of proofreading has, to our knowledge, remained unexplored. Proofreading in gene regulation is thus expected to increase the output dynamic range, which is favorable for signaling, but also to increase the noise, which is detrimental.

How can we formalize the trade-off between noise and dynamic range for gene regulatory schemes and find when proofreading is beneficial? In existing analyses of proofreading, the erroneous incorporation of the substrate leads to an error product that is *different* from the correct one [22,26]; in contrast, here the gene always expresses the *same* mRNA. What is important for signal transduction, however, is how well this expression correlates with the input signal, c. To quantify the regulatory power of the proofreading architecture, we computed the mutual

information, $I(c;m)$ [27], between the signal c and the mRNA expression level m, following previous applications of information theory to gene regulation [24,28]. The information depends not only on $P(m|c)$, which we compute from the master equation, but also on the *a priori* unknown distribution of input concentrations, $P(c)$; we therefore determined the input distribution $P^*(c)$ that maximizes information transmission, subject to a constraint on the average number of expressed mRNA, $\bar{m} = \int dc P(c) \sum_m m P(m|c)$. This constraint on the average number of mRNA was imposed to compare different regulatory architectures; otherwise, a higher average expression could yield higher information transmission, for trivial reasons. Such constrained information (capacity) maximization is a well-known problem in information theory that can be solved using the Blahut-Arimoto algorithm [29].

Figure 2(a) shows how the information transmission $I(m;c)$ through the promoter depends on the (inverse) reaction rate qd. We start by looking at the classic measure of proofreading performance, the "error fraction," i.e., the ratio of the mRNA expressed from state 2_{nc} due to noncognate TFs vs mRNA expressed from state 2_c due to cognate TFs. As qd is increased, the error fraction drops, with no clear optimum. In contrast, there exists an optimal q^*d at which the information is maximized—this is the point where proofreading is most effective, optimally trading off erroneous induction (here, suppressed by a factor of ~30 relative to no proofreading), noise in gene expression, and dynamic range at the output. In Fig. 2(b) we plot the noise in gene expression, as a function of the input concentration c for the optimal proofreading architecture and the nonproofreading limit. In both cases the noise has super-Poisson components due to the switching between promoter states, but this excess is substantially higher in the proofreading architecture, as expected.

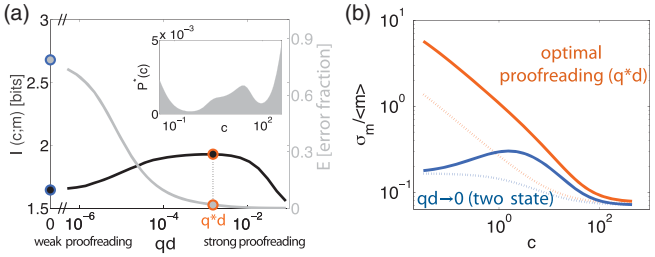

FIG. 2 (color online). (a) Maximal information transmission (left axis, black) and the error fraction (right axis, gray) as a function of the inverse irreversible reaction rate, qd. Increasing qd suppresses the error fraction, but only at the cost of increasing the gene expression noise, leading to a trade-off and an information-maximizing value of q^*d (orange). This maximum is reached robustly with input distributions that are close to optimal (inset; Ref. [30]). (b) Noise in gene expression, $\sigma_m/\langle m \rangle$, computed from the moments of $P(m|c)$, as a function of the dimensionless input concentration c, for the optimal proofreading (orange lines) and the two-state (blue lines) architectures. Dotted lines show the Poisson limit, $\sigma_m^2 = \langle m \rangle$, for comparison. In both cases, the average number of mRNA expressed is fixed to $\bar{m} = 100$.

While attractive, these results still depend on the particular rates chosen for the model in Fig. 1(b). Surprisingly, if we choose to compare the optimal proofreading scenario with the optimal nonproofreading one, the problem simplifies further. Given that the input TF concentration c varies over some limited dynamic range, $c \in [0, C_{max} = k_+ C/d]$, there should also exist an optimal setting for k_-^c: set too high, the cognate TFs will be extremely unlikely to occupy the promoter for any significant fraction of the time and induce the gene; set too low, the switching contribution to noise in gene expression will blow up. With k_-^c and q in the "correct initiation" pathway of Fig. 1(b) set by optimization, the remaining rates in the "erroneous initiation" pathway are fixed by the choice of crosstalk severity Λ. The remaining parameters regulating mRNA expression—the average mRNA count $\bar{m}$ and the rate r—do not change the results qualitatively. The mRNA expression rate r simply sets the maximal number of mRNA molecules at full expression in steady state (r/d); this influences the Poisson noise at the output, but does so equally for any regulatory architecture, proofreading or not. As long as r is large enough so that the average mRNA constraint $\bar{m}$ is achievable, the precise choice of these values is not crucial (we use $r/d = 200$, $\bar{m} = 100$, plausible for eukaryotic expression). In sum, we can compare how well the optimal proofreading architecture does compared to optimal nonproofreading architecture in terms of information transmission, as a function of two key parameters: the crosstalk severity, Λ, and the input dynamic range, C_{max}.

Figure 3(a) shows the advantage, in bits, of the optimal proofreading architecture relative to the optimal nonproofreading one. This "information plane," $I_{q*}(m;c) - I_{q=0}(m;c)$, is plotted as a function of Λ and C_{max}. In the limit $\Lambda \to 0$, the difference in performance goes to zero: there, optimization drives $q^* k_-^{nc,*} \gg 1$, but proofreading offers vanishing an advantage over the optimal two-state promoter architecture when noncognate binding is negligible. As Λ increases, proofreading becomes beneficial over the two-state architecture, and more so for higher values of C_{max}. Higher input concentrations $c \in [0, C_{max}]$ permit faster on rates, resulting in faster optimal off rates $k_-^{c,*}$ and faster optimal $1/q^*$. Generally, faster switching of promoter states in Fig. 1(b) means that promoter switching noise will be lower and thus the information higher (at fixed mean mRNA expression $\bar{m}$); in particular, optimization tends to minimize promoter switching noise by selecting the fastest $1/q$ that still admits error rejection; i.e., $q^* k_-^{nc,*} \sim 1$. At $\Lambda = \nu/\sigma \simeq 1$, the signaling capacity of the nonproofreading architecture collapses completely, with $I_{q=0}(c;m) \approx 0$ [32]. At this point optimal

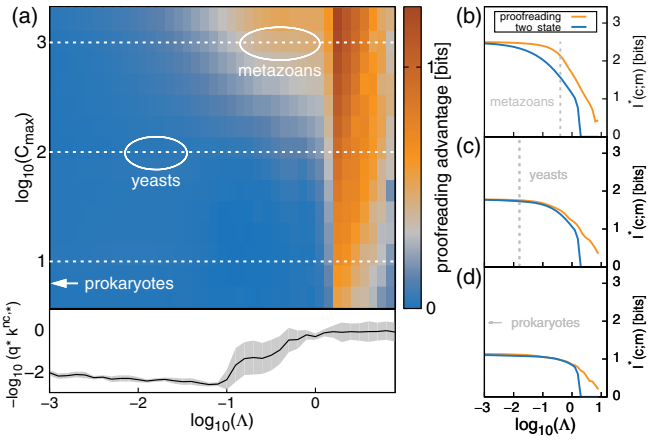

FIG. 3 (color online). (a) Information advantage (in bits, color scale) of optimal proofreading over optimal two-state architectures, as a function of crosstalk severity Λ and dynamic range of input TF concentration, $C_{\max}$. Typical values for prokaryotes, yeast, and metazoans are marked in white. (Lower inset) Optimal rates, $q^* k_-^{\mathrm{nc},*}$ (the black line = the average over $C_{\max}$, the gray shade = std), indicate a switch to the proofreading strategy. (b)–(d) Cuts through the information plane in (a) along the white dashed lines showing the collapse of the two-state performance as $\log_{10}(\Lambda) \to 0$ and a clear proofreading advantage for metazoan regulation.

proofreading architectures are affected, but they still generally maintain at least half of the capacity seen at $\Lambda = 0$; proofreading extends the performance of the gene regulation well into the $\Lambda > 0$ region before finally succumbing to crosstalk.

Where do different organisms lie in the information plane? Prokaryotes have on the order of $\nu \sim 100$ types of transcription factors, whose binding site motifs typically contain around 23 bits of sequence information [3], or a $16 k_B T$ binding energy difference between the cognate and noncognate sites [33], corresponding to $\sigma \sim 10^7$. The resulting crosstalk severity is low, $\Lambda \sim 10^{-5}$. For yeast, the typical sequence information is 14 bits ($10 k_B T$) [3], which gives $\Lambda \sim 0.01$ (for $\nu \sim 200$ [34]). For multicellular eukaryotes, the typical sequence information is 12 bits ($8 k_B T$), and the number of TF species varies between $\nu \approx 10^3$ (*C. elegans*) to $\nu \approx 2 \times 10^3$ (human) [35], putting Λ between 0.1 and 1. We can also estimate the dimensionless parameter $C_{\max} = k_+ C/d$. Assuming diffusion-limited binding of TFs to their binding sites, $k_+ C/d \approx 3DaN/R^3 d$, where $D \sim 1\ \mu\mathrm{m}^2/\mathrm{s}$ is the typical TF diffusion constant [35], $a \sim 3$ nm is the binding site size, $R = 3\ \mu$m (1 μm) is the radius of an eukaryotic nucleus (prokaryotic cell), and N is the typical copy number of TFs per nucleus ($N \sim 10$ for prokaryotes, 10^3 for yeast, 10^3–10^5 for eukaryotes). Typical mRNA lifetimes are 5–10 min in prokaryotes, 20–30 min in yeast, and > 1 hour in metazoans. This yields $C_{\max}$ of order 10 for prokaryotes, 10^2 for yeast cells, and $> 10^3$ for multicellular eukaryote cells. While these are very rough estimates, different kinds of cells clearly differ substantially in their location on the information plane of Fig. 3(a).

Taken together, these values suggest that crosstalk is acute for metazoans and that proofreading in gene regulation could provide a vast improvement over regulation at equilibrium, as in Fig. 3(b). In the Supplementary Material [36], we examine two possible molecular implementations of the abstract scheme outlined in Fig. 1(b): the first utilizes covalent modifications of the RNA polymerase II *C*-terminal domain (CTD) tail [48]; the second relies on histone modifications. How could these proofreading ideas be tested? Indirect evidence for kinetic schemes in regulation exists. Crystal structure of RNA polymerase II during early promoter clearance indicates that abortive initiation is a side product of "promoter proofreading" [49]. Experimentally documented interactions between histone tail modifiers, chromatin remodelers, and TFs appear to be consistent with kinetic proofreading [50]. Kinetic studies of gene activation by TF binding are inconsistent with equilibrium models [51]. Direct evidence showing that TF specificity is boosted by proofreading to reduce erroneous gene regulation is, however, lacking. Tests following [52] to measure ATP consumption per mRNA upon initiation due to cognate vs noncognate TFs appear possible *in vitro* for RNA polymerase II CTD modification mechanism, but they are difficult for histone-based mechanisms, which might be better tested indirectly using genetic perturbations [36].

While we cannot rule out the existence of a complex equilibrium scheme that reduces crosstalk in gene regulation sufficiently, this and our related work [16] suggest that equilibrium solutions, if they exist, are not simple. Here we advanced an alternative hypothetical mechanism, proofreading-based transcriptional regulation, to mitigate the crosstalk problem. Unlike most biophysical problems, where we clearly appreciate their out-of-equilibrium nature, transcriptional regulation has remained a textbook example of a nontrivial *equilibrium* molecular recognition process, likely due to the success of the equilibrium assumption in prokaryotes. Crosstalk considerations should motivate us to reexamine this assumption in eukaryotic regulation.

We thank T. R. Sokolowski and T. Friedlander for the helpful comments on the manuscript, and Erik van Nimwegen for suggesting that histone modification or remodeling might also constitute a proofreading mechanism candidate. G. R. acknowledges support from ERC Advanced Grant No. 250152.

[1] M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Press, New York, 2002).
[2] T. Kuhlman, Z. Zhang, M. H. Saier, Jr, and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **104**, 6043 (2007).

[3] Z. Wunderlich and L. A. Mirny, Trends Genet. **25**, 434 (2009).
[4] M. Tuğrul, T. Paixão, N. H. Barton, and G. Tkačik, PLoS Genet. **11**, e1005639 (2015).
[5] A. Sandelin *et al.*, Nucleic Acids Res. **32**, 91D (2004).
[6] X. Y. Li *et al.*, PLoS Biol. **6**, e27 (2008).
[7] S. Rockel, M. Geertz, K. Hens, B. Deplancke, and S. J. Maerkl, Nucleic Acids Res. **41**, e52 (2012).
[8] M. A. Shea and G. K. Ackers, J. Mol. Biol. **181**, 211 (1985).
[9] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Curr. Opin. Genet. Dev. **15**, 116 (2005).
[10] J. B. Kinney, A. Murugan, C. G. Callan, Jr., and E. C. Cox, Proc. Natl. Acad. Sci. U.S.A. **107**, 9158 (2010).
[11] H. Janssens, S. Hous, J. Jaeger, A. R. Kim, E. Myasnikova, D. Sharp, and J. Reinitz, Nat. Genet. **38**, 1159 (2006).
[12] X. He, M. A. H. Samee, C. Blatti, and S. Sinha, PLoS Comput. Biol. **6**, e1000935 (2010).
[13] W. D. Fakhouri, A. Ay, R. Sayal, J. Dresch, E. Dayringer, and D. N. Arnosti, Mol. Syst. Biol. **6**, 341 (2010).
[14] S. J. Maerkl and S. R. Quake, Science **315**, 233 (2007).
[15] L. Giorgetti, T. Siggers, G. Tiana, G. Caprara, S. Notarbartolo, T. Corona, M. Pasparakis, P. Milani, M. L. Bulyk, and G. Natoli, Mol. Cell **37**, 418 (2010).
[16] T. Friedlander, R. Prizak, C. G. Guet, N. H. Barton and G. Tkačik, arXiv:1506.06925.
[17] L. A. Mirny, Proc. Natl. Acad. Sci. U.S.A. **107**, 22534 (2010).
[18] A. L. Todeschini, A. Georges, and R. A. Veitia, Trends Genet. **30**, 211 (2014).
[19] W. McGinnis and R. Krumlauf, Cell **68**, 283 (1992).
[20] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, PLoS Biol. **4**, e309 (2006).
[21] S. C. Little, M. Tikhonov, and T. Gregor, Cell **154**, 789 (2013).
[22] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **71**, 4135 (1974).
[23] J. Peccoud and B. Ycart, Theor. Popul. Biol. **48**, 222 (1995).
[24] G. Rieckh and G. Tkačik, Biophys. J. **106**, 1194 (2014).
[25] G. Bel, B. Munsky, and I. Nemenman, Phys. Biol. **7**, 016003 (2009).
[26] Y. Savir and T. Tlusty, Cell **153**, 471 (2013).
[27] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Champaign, IL, 1949).
[28] A. M. Walczak and G. Tkačik, J. Phys. Condens. Matter **23**, 153102 (2011).
[29] R. E. Blahut, IEEE Trans. Inf. Theory **18**, 460 (1972).
[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.115.248101, which also includes Ref. [31], for further tests of robustness to perturbations in the input distribution.
[31] G. Tkačik, C. G. Callan, Jr., and W. Bialek, Phys. Rev. E **78**, 011910 (2008).
[32] This is independent of whether one modulates Λ by changing ν, as for Fig. 3(a), or by changing σ; although the optimal rates may take on different values, the information plane is essentially unchanged, irrespective of how Λ is modulated.
[33] U. Gerland, J. D. Moroz, and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **99**, 12015 (2002).
[34] R. Jothi, S. Balaji, A. Wuster, J. A. Grochow, J. Gsponer, T. M. Przytycka, L. Aravind, and M. Madan Babu, Mol. Syst. Biol. **5**, 294 (2009).
[35] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, Nucleic Acids Res. **38**, D750 (2009).
[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.115.248101, which includes Refs. [37–47], for RNA polymerase II and histone-based implementations of proofreading.
[37] S. Buratowski, Mol. Cell **36**, 541 (2009).
[38] S. Egloff, M. Dienstbier, and S. Murphy, Trends Genet. **28**, 333 (2012).
[39] G. Prelich, Eukaryotic Cell **1**, 153 (2002).
[40] H. Cho, T.-K. Kim, H. Mancebo, W. S. Lane, O. Flores, and D. Reinberg, Genes Dev. **13**, 1540 (1999).
[41] J. W. Stiller and B. D. Hall, Proc. Natl. Acad. Sci. U.S.A. **99**, 6091 (2002).
[42] R. J. Taft, M. Pheasant, and J. S. Mattick, BioEssays **29**, 288 (2007).
[43] E. van Nimwegen, Trends Genet. **19**, 479 (2003).
[44] M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein, and S. A. Teichmann, Curr. Opin. Struct. Biol. **14**, 283 (2004).
[45] S. Venkatesh and J. L. Workman, Nat. Rev. Mol. Cell Biol. **16**, 178 (2015).
[46] C. M. Weber, S. Ramachandran, and S. Henikoff, Mol. Cell **53**, 819 (2014).
[47] A. Nock, J. M. Ascano, M. J. Barrero, and S. Malik, Mol. Cell **48**, 837 (2012).
[48] S. Egloff and S. Murphy, Trends Genet. **24**, 280 (2008).
[49] X. Liu, D. A. Bushnell, D.-A. Silva, X. Huang, and R. D. Kornberg, Science **333**, 633 (2011).
[50] R. Blossey and H. Schiessel, Biophys. J. **101**, L30 (2011).
[51] J. Chen *et al.*, Cell **156**, 1274 (2014).
[52] J. J. Hopfield, T. Yamane, V. Yue, and S. M. Coutts, Proc. Natl. Acad. Sci. U.S.A. **73**, 1164 (1976).