

**Optimizing information flow in small genetic networks**Gašper Tkačik,<sup>1,\*</sup> Aleksandra M. Walczak,<sup>2,†</sup> and William Bialek<sup>2,3,‡</sup><sup>1</sup>*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6396, USA*<sup>2</sup>*Joseph Henry Laboratories of Physics, Lewis-Sigler Institute for Integrative Genomics, and Princeton Center for Theoretical Science, Princeton University, Princeton, New Jersey 08544, USA*<sup>3</sup>*Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10065, USA*

(Received 25 March 2009; revised manuscript received 12 July 2009; published 29 September 2009)

In order to survive, reproduce, and (in multicellular organisms) differentiate, cells must control the concentrations of the myriad different proteins that are encoded in the genome. The precision of this control is limited by the inevitable randomness of individual molecular events. Here we explore how cells can maximize their control power in the presence of these physical limits; formally, we solve the theoretical problem of maximizing the information transferred from inputs to outputs when the number of available molecules is held fixed. We start with the simplest version of the problem, in which a single transcription factor protein controls the readout of one or more genes by binding to DNA. We further simplify by assuming that this regulatory network operates in steady state, that the noise is small relative to the available dynamic range, and that the target genes do not interact. Even in this simple limit, we find a surprisingly rich set of optimal solutions. Importantly, for each locally optimal regulatory network, all parameters are determined once the physical constraints on the number of available molecules are specified. Although we are solving an oversimplified version of the problem facing real cells, we see parallels between the structure of these optimal solutions and the behavior of actual genetic regulatory networks. Subsequent papers will discuss more complete versions of the problem.

DOI: [10.1103/PhysRevE.80.031920](https://doi.org/10.1103/PhysRevE.80.031920)

PACS number(s): 87.10.Vg, 87.15.R–, 87.18.Tt, 87.85.Ng

**I. INTRODUCTION**

Much of the everyday business of organisms involves the transmission and processing of information. On our human scale, the familiar examples involve the signals taken in through our sense organs [1]. On a cellular scale, information flows from receptors on the cell surface into the cell, modulating biochemical events and ultimately controlling gene expression [2]. In the course of development in multicellular organisms, individual cells acquire information about their location in the embryo by responding to particular “morphogen” molecules whose concentration varies along the main axes of the embryo [3,4]. In all these examples, information of interest to the organism ultimately is represented by events at the molecular level, whether the molecules are transcription factors regulating gene expression or ion channels controlling electrical signals in the brain. This representation is limited by fundamental physical principles: individual molecular events are stochastic, so that with any finite number of molecules there is a limit to the precision with which small signals can be discriminated reliably and there is a limit to the overall dynamic range of the signals. Our goal in this paper (and its sequel) is to explore these limits to information transmission in the context of small genetic control circuits.

The outputs of genetic control circuits are protein molecules that are synthesized by the cell from messenger RNA (mRNA), which in turn is transcribed from the DNA template. The inputs often are protein molecules as well, “tran-

scription factors” that bind to the DNA and regulate the synthesis of the mRNA. In the last decade, a number of experiments have mapped the input/output relations of these regulatory elements and characterized their noise, that is, the fluctuations in the output protein concentration when the inputs are held fixed [5–17]. In parallel, a number of theoretical papers have tried to understand the origins of this noise, which ultimately reflects the random behavior of individual molecules along the path from input to output—the arrival of transcription factors at their targets along the DNA, the initiation of transcription and the degradation of mRNA, and the initiation of protein synthesis and the degradation of the output proteins [18–29]. While open questions remain, it seems fair to say that we have a physical picture of the noise in genetic control that we can use to ask questions about the overall function and design of these systems.

The ability of any system to transmit information is determined not just by input/output relations and noise levels, but also by the distribution of inputs; maximal information transmission requires a matching between the intrinsic properties of the system and the input statistics [30,31]. In the context of sensory information processing, these matching conditions have been explored almost since the inception of information theory [32–35]. In particular, because the distribution of sensory inputs varies with time, optimal information transmission requires that the input/output relation track or adapt to these variations and this theoretical prediction has led to a much richer view of adaptation in the neural code [36–40]. There are analogous matching conditions for genetic regulatory elements and these conditions provide parameter free predictions about the behavior of the system based on the idea that cells are trying to transmit the maximum amount of information [41]. Comparison with recent experiments has been encouraging [42].

\*gkacik@sas.upenn.edu

†awalczak@princeton.edu

‡wbialek@princeton.edu

In this paper we go beyond the matching conditions to ask how cells can adjust the input/output relations of genetic regulatory elements so as to maximize the information that is transmitted through these systems. Absent any constraints, the answer will always be to make more molecules, since this reduces the effective noise level, so we consider the problem of maximizing information transmission with a fixed mean or maximum number of molecules at both the input and the output. In this sense we are asking how cells can extract the maximum control power, measured in bits, from a given number of molecules, thus optimizing functionality under clear physical constraints. In general this problem is very difficult, so we start here with the simplest case of a single input transcription factor that controls (potentially) many genes, but there is no interaction among these outputs. Further, we focus on a limit (small noise) where some analytic progress is possible. We will see that, even in this case, the optimal solutions have an interesting structure, which emerges as a result of the interplay between noise sources at the input and the output of the regulatory elements. For other approaches to the optimization of information transmission in biochemical and genetic networks, see Refs. [43–45].

Optimization of information transmission is a concise, abstract principle grounded in the physics of the molecular interactions that underlie biological function. It would be attractive if we could derive the behavior of biological systems from such a principle rather than taking the myriad parameters of these systems simply as quantities that must be fit to data. It is not at all clear, however, that such a general principle should apply to real biological systems. Indeed, it is possible that solutions to our optimization problem are far from plausible in comparison with what we find in real cells. Thus, our most important result is that the parameters which we derive are reasonable in relation to experiment. While a realistic comparison requires us to solve the optimization problem in a fully interacting system, even in the simpler problem discussed here we can see the outlines of a theory for real genetic networks. Subsequent papers will address the full, interacting version of the problem.

## II. FORMULATING THE PROBLEM

A gene regulatory element translates the concentration of input molecules  $\mathcal{I}$  into output molecules  $\mathcal{O}$ . We would like to measure, quantitatively, how effectively changes in the input serve to control the output. If we make many observations on the state of the cell, we will see that inputs and outputs are drawn from a joint distribution  $p(\mathcal{I}, \mathcal{O})$  and our measure of control power should be a functional of this distribution. In his classic work, Shannon showed that there is only one such measure of control power which obeys certain plausible constraints and this is the mutual information between  $\mathcal{I}$  and  $\mathcal{O}$  [30,46].

To be concrete, we consider a set of genes,  $i = 1, 2, \dots, M$ , that all are controlled by a single transcription factor. Let the concentration of the transcription factor be  $c$  and let the levels of protein expressed from each gene be  $g_i$ ; below we discuss the units and normalization of these quantities. Thus, the input  $\mathcal{I} \equiv c$  and the output  $\mathcal{O} \equiv \{g_i\}$ . In prin-

ciple these quantities all depend on time. We choose to focus here on the steady-state problem, where we assume that the output expression levels reach their equilibrium values before the input transcription factor concentrations change.

We view the steady-state approximation not necessarily as an accurate model of the dynamics in real cells, but as a useful starting point, and already the steady-state problem has a rich structure. In particular, as we will see, in this limit we have analytic control over the role of nonlinearities in the input/output relation describing the function of the different regulatory elements in our network. In contrast, most approaches to information transmission by dynamic signals are limited to the regime of linear response; see, for example, Ref. [45]. Although we are focused here on information transmission in genetic circuits, it is interesting that the same dichotomy—nonlinear analyses of static networks and dynamic analyses of linear networks—also exists in the literature on information transmission in neural networks [34,35].

To specify the joint distribution of inputs and outputs, it is convenient to think that the transcription factor concentration is being chosen out of a probability distribution  $P_{TF}(c)$  and then the target genes respond with expression levels chosen out of the conditional distribution  $P(\{g_i\}|c)$ . In general, the mutual information between the set of expression levels  $\{g_i\}$  and the input  $c$  is given by [30,31]

$$I(\{g_i\}; c) = \int dc \int d^M g P(c, \{g_i\}) \log_2 \left[ \frac{P(c, \{g_i\})}{P_{TF}(c) P(\{g_i\})} \right], \quad (1)$$

where the overall distribution of expression levels is given by

$$P(\{g_i\}) = \int dc P_{TF}(c) P(\{g_i\}|c). \quad (2)$$

Shannon's uniqueness theorem of course leaves open a choice of units and here we make the conventional choice of bits, hence, the logarithm is base 2.

We will approach the problem of optimizing information transmission in two steps. First, we will adjust the distribution  $P_{TF}(c)$  to take best advantage of the input/output relations and then we will adjust the input/output relations themselves. Even the first step is difficult in general, so we start by focusing on the limit in which noise is small.

### A. Information in the small noise limit

As noted in Sec. I, we will confine our attention in this paper to the case where each gene responds independently to its inputs and there are no interactions among the output genes; we point toward generalizations in Sec. V below and return to the more general problem in subsequent papers. The absence of interactions means that the conditional distribution of expression levels must factorize,  $P(\{g_i\}|c) = \prod_{i=1}^M P_i(g_i|c)$ . Further, we assume that the noise in expression levels is Gaussian. Then we have [47]

$$P(\{g_i\}|c) = \exp \left\{ -\frac{M}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^M \ln[\sigma_i^2(c)] - \frac{1}{2} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} [g_i - \bar{g}_i(c)]^2 \right\}. \quad (3)$$

The input/output relation of each gene is defined by the mean  $\bar{g}_i(c)$ , while  $\sigma_i^2$  measures the variance of the fluctuations or noise in the expression levels at fixed input,

$$\sigma_i^2(c) = \langle [g_i - \bar{g}_i(c)]^2 \rangle. \quad (4)$$

In the limit that the noise levels  $\sigma_i$  are small, we can develop a systematic expansion of the information  $I(\{g_i\}; c)$  generalizing the approach of Refs. [41,42]. The key idea is that, in the small noise limit, observation of the output expression levels  $\{g_i\}$  should be sufficient to determine the input concentration  $c$  with relatively high accuracy; further, we expect that errors in this estimation process would be well approximated as Gaussian. Formally, this means that we should have

$$P(c|\{g_i\}) \approx \frac{1}{\sqrt{2\pi\sigma_c^2(\{g_i\})}} \exp \left[ -\frac{[c - c^*(\{g_i\})]^2}{2\sigma_c^2(\{g_i\})} \right], \quad (5)$$

where  $c^*(\{g_i\})$  is the most likely value of  $c$  given the outputs and  $\sigma_c^2(\{g_i\})$  is the variance of the true value around this estimate. We can use this expression to calculate the information by writing  $I(\{g_i\}; c)$  as the difference between two entropies:

$$I(\{g_i\}; c) = - \int dc P_{TF}(c) \log_2 P_{TF}(c) - \int d^M g P(\{g_i\}) \times \left[ - \int dc P(c|\{g_i\}) \log_2 P(c|\{g_i\}) \right] \quad (6)$$

$$= - \int dc P_{TF}(c) \log_2 P_{TF}(c) - \frac{1}{2} \int d^M g P(\{g_i\}) \log_2 [2\pi e \sigma_c^2(\{g_i\})]. \quad (7)$$

Intuitively, the first term is the entropy of inputs, which sets an absolute maximum on the amount of information that can be transmitted [48]; the second term is (minus) the entropy of the input given the output or the “equivocation” [30] that results from noise in the mapping from inputs to outputs. To complete the calculation we need an expression for this effective noise level  $\sigma_c$ .

Using Bayes’ rule, we have

$$P(c|\{g_i\}) = \frac{P(\{g_i\}|c) P_{TF}(c)}{P(\{g_i\})} \quad (8)$$

$$= \frac{1}{Z(\{g_i\})} \exp[-F(c, \{g_i\})], \quad (9)$$

where

$$F(c, \{g_i\}) = - \ln P_{TF}(c) + \frac{1}{2} \sum_{i=1}^M \ln[\sigma_i^2(c)] + \frac{1}{2} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} [g_i - \bar{g}_i(c)]^2. \quad (10)$$

Now it is clear that  $c^*(\{g_i\})$  and  $\sigma_c(\{g_i\})$  are defined by

$$0 = \left. \frac{\partial F(c, \{g_i\})}{\partial c} \right|_{c=c^*(\{g_i\})}, \quad (11)$$

$$\frac{1}{\sigma_c^2(\{g_i\})} = \left. \frac{\partial^2 F(c, \{g_i\})}{\partial c^2} \right|_{c=c^*(\{g_i\})}. \quad (12)$$

The leading term at small  $\sigma_i$  is then given by

$$\frac{1}{\sigma_c^2(\{g_i\})} = \sum_{i=1}^M \frac{1}{\sigma_i^2} \left( \frac{d\bar{g}_i(c)}{dc} \right)^2 \Big|_{c=c^*(\{g_i\})}. \quad (13)$$

Finally, we note that, in the small noise limit, averages over all the expression levels can be approximated by an integral along the trajectory of mean expression levels with an appropriate Jacobian. More precisely,

$$\int d^M g P(\{g_i\}) [\dots] \approx \int dc P_{TF}(c) \prod_{i=1}^M \delta(g_i - \bar{g}_i(c)) [\dots]. \quad (14)$$

Putting all these terms together, we have

$$I(\{g_i\}; c) = - \int dc P_{TF}(c) \log_2 P_{TF}(c) + \frac{1}{2} \int dc P_{TF}(c) \log_2 \left[ \frac{1}{2\pi e} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} \left( \frac{d\bar{g}_i(c)}{dc} \right)^2 \right]. \quad (15)$$

The small noise approximation is not just a theorist’s convenience. A variety of experiments show that fluctuations in gene expression level can be 10%–25% of the mean [5,9,10,13,14,17]. As noted above, maximizing information transmission requires matching the distribution of input signals to the structure of the input/output relations and noise, and in applying these conditions to a real regulatory element in the fruit fly embryo it was shown that the (analytically accessible) small noise approximation gives results which are in semiquantitative agreement with the (numerical) exact solutions [42]. Thus, although it would be interesting to explore the quantitative deviations from the small noise limit, we believe that this approximation is a good guide to the structure of the full problem.

To proceed, Eq. (15) for the information in the small noise limit instructs us to compute the mean response  $\bar{g}_i(c)$  and the noise  $\sigma_i(c)$  for every regulated gene. Since the properties of noise in gene expression determine to a large extent the structure of optimal solutions, we present in Sec. II B a detailed description of these noise sources. In Sec. II C we then introduce the “cost of coding” measured by the number of signaling molecules that the cell has to pay to transmit the

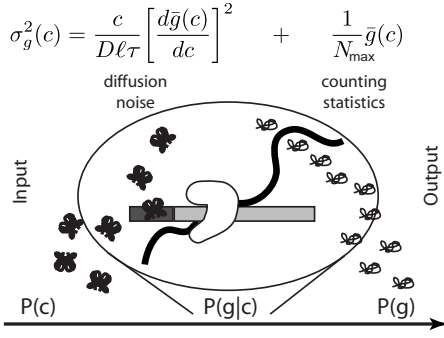


FIG. 1. Input proteins at concentration  $c$  act as transcription factors for the expression of output proteins,  $g$ . The diffusive noise in transcription factor concentration and the shot noise at the output both contribute to stochastic gene expression. The regulation process is described using a conditional probability distribution of the output knowing the input,  $P(g|c)$ , which can be modeled as a Gaussian process with a variance  $\sigma_g^2(c)$ . In this paper we consider the case of multiple output genes  $\{g_i\}$ ,  $i=1, \dots, M$ , each of which is independently regulated by the process illustrated here with the corresponding noise  $\sigma_i^2$ .

information reliably. Finally, we look for optimal solutions in Sec. III.

### B. Input/output relations and noise

Transcription factors act by binding to DNA near the point at which the “reading” of a gene begins, and either enhancing or inhibiting the process of transcription into mRNA. In bacteria, a simple geometrical view of this process seems close to correct and one can try to make a detailed model of the energies for binding of the transcription factor(s) and the interaction of these bound factors with the transcriptional apparatus, RNA polymerase in particular [49,50]. For eukaryotes the physical picture is less clear, so we proceed phenomenologically. If binding of the transcription factor activates the expression of gene  $i$ , we write

$$\bar{g}_i(c) = \frac{c^{n_i}}{c^{n_i} + K_1^{n_i}}, \quad (16)$$

and similarly if the transcription factor represses expression we write

$$\bar{g}_i(c) = \frac{K_1^{n_i}}{c^{n_i} + K_1^{n_i}}. \quad (17)$$

These are smooth monotonic functions that interpolate between roughly linear response ( $n=1$  and large  $K$ ) and steep, switchlike behavior ( $n \rightarrow \infty$ ) at some threshold concentration ( $c=K$ ). Such “Hill functions” often are used to describe the cooperative binding of  $n$  molecules to their target sites [51], with  $F = -k_B T \ln K$  the free energy of binding per molecule, and this is a useful intuition even if it is not correct in detail.

To complete our formulation of the problem we need to understand the noise or fluctuations in expression level at fixed inputs as summarized by variances  $\sigma_i^2$ . There are several contributions to the variance, which we can divide into two broad categories, as in Fig. 1.

The transcription of mRNA and its translation into protein can be thought of as the “output” side of the regulatory apparatus. Ultimately these processes are composed of individual molecular events, and so there should be shot noise from the inherent randomness of these events. This suggests that there will be an output noise variance proportional to the mean,  $\sigma_{i,\text{out}}^2 \propto \bar{g}_i$ .

The arrival of transcription factor molecules can be thought of as the “input” side of the apparatus and again there should be noise associated with the randomness in this arrival. This noise is equivalent to a fluctuation in the input concentration itself; the variance in concentration should again be proportional to the mean and the impact of this noise needs to be propagated through the input/output relation so that  $\sigma_{i,\text{in}}^2 \propto c(d\bar{g}_i/dc)^2$ .

Putting together the input and output noise, we have

$$\sigma_i^2(c) = a\bar{g}_i(c) + bc \left( \frac{d\bar{g}_i(c)}{dc} \right)^2, \quad (18)$$

where  $a$  and  $b$  are constants. Comparing this intuitive estimate to more detailed calculations [21,29] allows us to interpret these constants. If  $\bar{g}_i$  is normalized so that its maximum value is 1, then  $a=1/N_{\text{max}}$ , where  $N_{\text{max}}$  is the maximum number of independent molecules that are made from gene  $i$ . If, for example, each mRNA molecule generates many proteins during its lifetime, then (if the synthesis of mRNA is limited by a single kinetic step)  $N_{\text{max}}$  is the maximum number of mRNAs, as discussed in Refs. [20,22,29].

The shot noise in the arrival of transcription factors at their targets ultimately arises from diffusion of these molecules. Analysis of the coupling between diffusion and the events that occur at the binding site [21,26,28] shows that the total input noise has both a term  $\propto c(d\bar{g}_i/dc)^2$  and additional terms that can be made small by adjusting the parameters describing kinetics of steps that occur after the molecules arrive at their target; here we assume that nature chooses parameters which make these nonfundamental noise sources negligible [52]. In the remaining term, we have  $b \sim 1/(D\ell\tau)$ , where  $D$  is the diffusion constant of the transcription factor,  $\ell$  is the size of its target on the DNA, and  $\tau$  is the time over which signals are integrated in establishing the steady state.

With the (semi)microscopic interpretation of the parameters, we can write

$$\sigma_i^2(c) = \frac{1}{N_{\text{max}}} \left[ \bar{g}_i(c) + cc_0 \left( \frac{d\bar{g}_i(c)}{dc} \right)^2 \right], \quad (19)$$

where there is a natural scale of concentration,

$$c_0 = \frac{N_{\text{max}}}{D\ell\tau}. \quad (20)$$

To get a rough feeling for this scale, we note that diffusion constants for proteins in the cytoplasm are  $\sim \mu\text{m}^2/\text{s}$  [16,53–55], target sizes are measured in nanometers, and integration times are minutes or hundreds of seconds (although there are few direct measurements). The maximum number of independent molecules depends on the character of the target genes. In many cases of interest, these are also tran-

scription factors, in which case a number of experiments suggest that  $N_{\max} \sim 10-100$  [12,22,29]. Putting these numbers together, we have  $c_0 \sim 10-100/(\mu\text{m})^3$  or  $\sim 15-150$  nM, although this (obviously) is just an order of magnitude estimate.

To summarize, two rather general forms of noise limit the information transmission in genetic regulatory networks. Both combine additively and ultimately trace their origin to a finite (and possibly small) number of signaling molecules. The input noise is caused by a small concentration of transcription factor molecules and its effect on the regulated gene is additionally modulated by the input-output relation. The output noise is caused by the small number of gene products and this noise is simply proportional to the mean. It is reasonable to believe that the strengths of these two noise sources, in appropriate units, will be of comparable magnitude. Since the organism has to pay a metabolic price to reduce either noise source, it would be wasting resources if it were to lower the strength of one source alone far below the limiting effect of the other.

### C. Constraining means or maxima

To proceed, we need to decide how the problem of maximizing information transmission will be constrained. One possibility is that we fix the maximum number of molecules at the input and the output. The constraint on the output can be implemented by measuring the expression levels in units such that the largest values of the mean expression levels  $\bar{g}_i$  are all equal to 1 [56]. On the input side, we restrict the range of  $c$  to be  $c \in [0, c_{\max}]$ . With this normalization and limits on the  $c$  integrals, we can maximize  $I(\{g_i\}; c)$  directly by varying the distribution of inputs, adding only a Lagrange multiplier to fix the normalization of  $P_{TF}(c)$ ,

$$\frac{\delta}{\delta P_{TF}(c)} \left[ I(\{g_i\}; c) - \lambda \int dc P_{TF}(c) \right] = 0. \quad (21)$$

As discussed in Ref. [42], the solution to the variational problem defined in Eq. (21) is

$$P_{TF}^*(c) = \frac{1}{Z_1 \sqrt{2\pi e} \sigma_c} \quad (22)$$

$$= \frac{1}{Z_1} \left[ \frac{1}{2\pi e} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} \left( \frac{d\bar{g}_i(c)}{dc} \right)^2 \right]^{1/2}, \quad (23)$$

where the normalization constant  $Z_1$  is given by

$$Z_1 = \int_0^{c_{\max}} dc \left[ \frac{1}{2\pi e} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} \left( \frac{d\bar{g}_i(c)}{dc} \right)^2 \right]^{1/2}. \quad (24)$$

The information transmission with this optimal choice of  $P_{TF}(c)$  takes a simple form,

$$I_1^* = \log_2 Z_1. \quad (25)$$

The expression for  $Z_1$ , and hence the optimal information transmission, has a simple geometric interpretation. As the concentration of the input transcription factor varies, the out-

put moves, on average, along a trajectory in the  $M$ -dimensional space of expression levels; this trajectory is defined by  $\{\bar{g}_i(c)\}$ . Nearby points along this trajectory cannot really be distinguished because of noise; the information transmission should be related to the number of distinguishable points. If the noise level was the same everywhere, this count of distinguishable states would be just the length of the trajectory in units where the standard deviation of the output fluctuations, projected along the trajectory, is 1. Since the noise is not uniform, we should introduce the local noise level into our metric for measuring distances in the space of expression levels and this is exactly what we see in Eq. (24). Thus, we can think of the optimal information transmission as being determined by the length of the path in expression space that the network traces as the input concentration varies, where length is measured with a metric determined by the noise level.

This information capacity still depends upon the input/output relations and the noise levels, so we have a second layer of optimization that we can perform. Before doing this, however, we consider another formulation of the constraints.

As an alternative to fixing the maximum concentration of input transcription factor molecules, we consider fixing the mean concentration. To do this, we introduce, as usual, a second Lagrange multiplier  $\alpha$ , so that our optimization problem becomes

$$\frac{\delta}{\delta P_{TF}(c)} \left[ I(\{g_i\}; c) - \lambda \int dc P_{TF}(c) - \alpha \int dc P_{TF}(c) c \right] = 0. \quad (26)$$

Notice that we can also think of this as maximizing information transmission in the presence of some fixed cost per input molecule.

Solving Eq. (26) for the distribution of inputs,  $P_{TF}(c)$ , we find

$$P_{TF}^*(c) = \frac{1}{Z_2} \left[ \frac{1}{2\pi e} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} \left( \frac{d\bar{g}_i(c)}{dc} \right)^2 \right]^{1/2} e^{-\alpha c}, \quad (27)$$

where

$$Z_2 = \int_0^{\infty} dc \left[ \frac{1}{2\pi e} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} \left( \frac{d\bar{g}_i(c)}{dc} \right)^2 \right]^{1/2} e^{-\alpha c}. \quad (28)$$

As usual in such problems we need to adjust the Lagrange multipliers to match the constraints, which is equivalent to solving

$$-\frac{\partial \ln Z_2}{\partial \alpha} = \langle c \rangle. \quad (29)$$

The optimal information transmission in this case is

$$I_2^* = \log_2 Z_2 + \alpha \langle c \rangle. \quad (30)$$

One might think that, for symmetry's sake, we should consider a formulation in which the mean number of output molecules also is constrained. There is some subtlety to this, since if we know the input/output functions,  $\{\bar{g}_i(c)\}$ , and the distribution of inputs,  $P_{TF}(c)$ , then the mean output levels are

determined. Thus, it is not obvious that we have the freedom to adjust the mean output levels. We return to this point in Sec. III C.

### III. ONE INPUT, ONE OUTPUT

To get a feeling for the structure of our optimization problem, we consider the case where the transcription factor regulates the expression level of just one gene. If we constrain the maximum concentrations at the input and output, then the information capacity is set by  $I = \log_2 Z_1$  [Eq. (25)]; substituting our explicit expression for the noise [Eq. (19)] we have

$$Z_1 = \int_0^{c_{\max}} dc \left[ \frac{N_{\max}}{2\pi e} \frac{(d\bar{g}(c)/dc)^2}{\bar{g}(c) + c_0 c [d\bar{g}(c)/dc]^2} \right]^{1/2}. \quad (31)$$

The first point to note is that if the natural scale of concentration,  $c_0$ , is either very large or very small, then the optimization problem loses all of its structure. Specifically, in these two limits we have

$$Z_1(c_0 \rightarrow \infty) = \left[ \frac{D\ell\tau}{2\pi e} \right]^{1/2} \int_0^{c_{\max}} \frac{dc}{\sqrt{c}}, \quad (32)$$

$$= \left[ \frac{2D\ell\tau c_{\max}}{\pi e} \right]^{1/2}, \quad (33)$$

and

$$Z_1(c_0 \rightarrow 0) = \left[ \frac{N_{\max}}{2\pi e} \right]^{1/2} \int_0^{c_{\max}} \frac{dc}{\sqrt{\bar{g}(c)}} \left| \frac{d\bar{g}(c)}{dc} \right|, \quad (34)$$

$$= \left[ \frac{2N_{\max}}{\pi e} \right]^{1/2} |\sqrt{\bar{g}(c_{\max})} - \sqrt{\bar{g}(0)}|. \quad (35)$$

In both cases, the magnitude of the information capacity becomes independent of the shape of the input/output relation  $\bar{g}(c)$ . Thus, the possibility that real input/output relations are determined by the optimization of information transmission depends on the scale  $c_0$  being comparable to the range of concentrations actually used in real cells. Although we have only a rough estimate of  $c_0 \sim 15\text{--}150$  nM, Table I shows that this is the case.

#### A. Numerical results with $c_{\max}$

To proceed, we choose  $c_0$  as the unit of concentration, so that

$$Z_1 = \left[ \frac{N_{\max}}{2\pi e} \right]^{1/2} \tilde{Z}_1, \quad (36)$$

$$\tilde{Z}_1(K/c_0, n; C) = \int_0^C dx \left[ \frac{(d\bar{g}(x)/dx)^2}{\bar{g}(x) + x[d\bar{g}(x)/dx]^2} \right]^{1/2}, \quad (37)$$

where  $C = c_{\max}/c_0$  and

TABLE I. Concentration scales for transcription factors. We collect absolute concentration measurements on transcription factors from several different systems, sometimes indicating the maximum observed concentration and in other cases the concentration that achieves half-maximal activation or repression (midpoint). Bcd is the bicoid protein, a transcription factor involved in early embryonic pattern formation; GAGA is a transcription factor in *Drosophila*; crp is a transcription factor that acts on a wide range of metabolic genes in bacteria; *lac* is the well studied operon that encodes proteins needed for lactose metabolism in *E. coli*; lac is the transcription factor that represses expression of the *lac* operon; OR1–3 are binding sites for the lac repressor.

Concentration	Scale	System	Ref.
$55 \pm 10$ nM	Midpoint	$\lambda$ repressor in <i>E. coli</i>	[10]
$55 \pm 3$ nM	Maximum	Bcd in <i>Drosophila</i> embryo	[17]
$5.3 \pm 0.7$ nM	Midpoint	GAGA	[57]
$\sim 5$ nM	Midpoint	crp to <i>lac</i> site	[50]
$\sim 0.2$ nM	Midpoint	lac to OR1	[50,58]
$\sim 3$ nM	Midpoint	lac to OR2	[50,58]
$\sim 110$ nM	Midpoint	lac to OR3	[50,58]
$22 \pm 3$ nM	Midpoint	lac to OR1 <i>in vitro</i>	[59]

$$\bar{g}(x) = \frac{x^n}{(K/c_0)^n + x^n} \quad (38)$$

in the case of an activator. It now is straightforward to explore, numerically, the function  $\tilde{Z}_1$ . An example, with  $c_{\max}/c_0 = 1$ , is shown in Fig. 2.

We see that, with  $c_{\max} = c_0$ , there is a well defined but broad optimum of the information transmission as a function

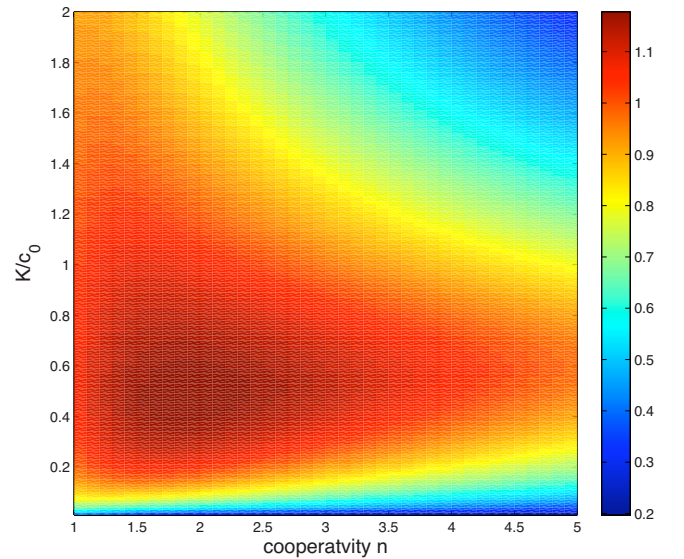


FIG. 2. (Color online) Information capacity for one (activator) input and one output. The information is  $I = \log_2 \tilde{Z}_1 + A$ , with  $A$  independent of the parameters; the map shows  $\tilde{Z}_1$  as computed from Eq. (37), here with  $C \equiv c_{\max}/c_0 = 1$ . We see that there is a broad optimum with cooperativity  $n_{\text{opt}} = 1.86$  and  $K_{\text{opt}} = 0.48c_0 = 0.48c_{\max}$ .

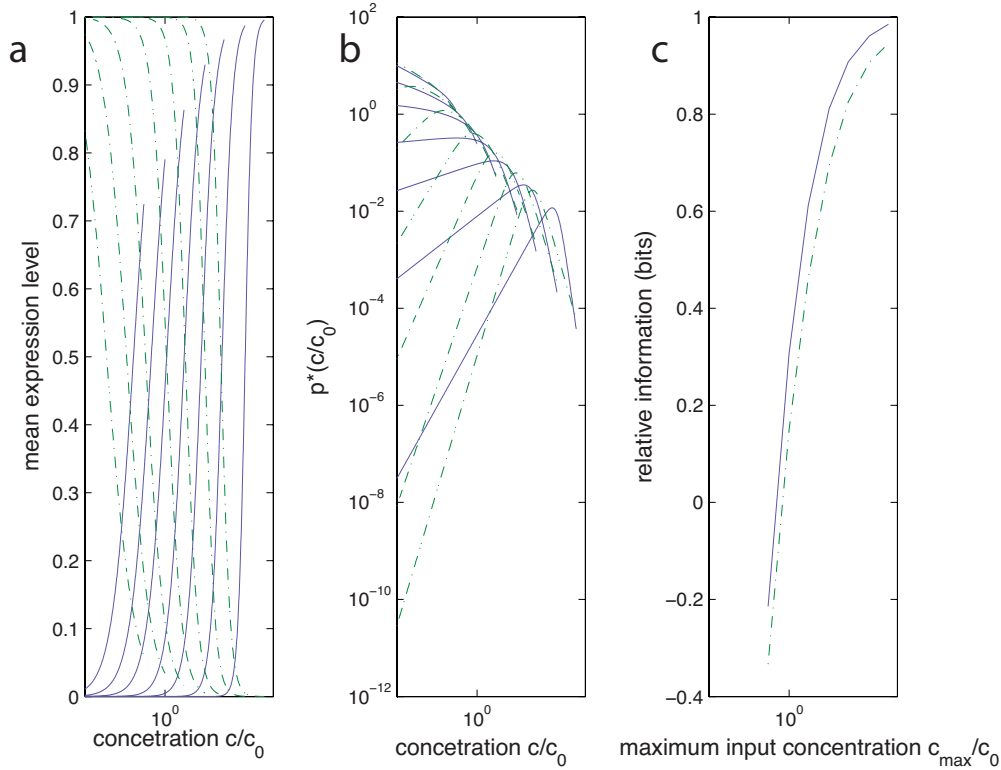


FIG. 3. (Color online) The optimal solutions for one gene controlled by one transcription factor. The optimization of information transmission in the small noise limit depends on only one parameter, which we take here as the maximum concentration of the input molecules, measured in units determined by the noise itself [ $c_0$  from Eq. (20)]. Panel (a) shows the optimal input/output relations with  $c_{\max}/c_0=0.3, 1, 3, 10, 30, 100, 300$ ; activators shown in blue (solid line), repressors in green (dashed line). Although the input/output relation is defined for all  $c$ , we show here only the part of the dynamic range that is accessed when  $0 < c < c_{\max}$ . Panel (b) shows the optimal distributions,  $P_{TF}^*(c)$ , for each of these solutions. Panel (c) plots  $\log_2 \tilde{Z}_1$  for these optimal solutions as a function of  $c_{\max}/c_0$ . Up to an additive constant, this is the optimal information capacity, in bits.

of the parameters  $K$  and  $n$  describing the input/output relation. Maximum information transmission occurs at modest levels of cooperativity ( $n \approx 2$ ) and with the midpoint of the input/output relation near the midpoint of the available dynamic range of input concentrations ( $K \approx c_{\max}/2$ ).

Optimal solutions for activators and repressors have qualitatively similar behaviors, with the optimal parameters  $K_{\text{opt}}$  and  $n_{\text{opt}}$  both increasing as  $c_{\max}$  increases [Fig. 3(a)]. Interestingly, at the same value of  $c_{\max}$ , the optimal repressors make fuller use of the dynamic range of outputs. The information capacity itself, however, is almost identical for activators and repressors across a wide range of  $c_{\max}$  [Fig. 3(c)]. This is important, because it shows that our optimization problem, even in this simplest form, can have multiple nearly degenerate solutions. We also see that increases in  $c_{\max}$  far beyond  $c_0$  produce a rapidly saturating information capacity, as expected from Eq. (35). Therefore, although increasing the dynamic range always results in an increase in capacity, the advantage in terms of information capacity gained by the cell being able to use input concentration regimes much larger than  $c_0$  is quite small.

### B. Some analytic results

Although the numerical results are straightforward, we would like to have some intuition about these optimal solu-

tions from analytic approximations. Our basic problem is to do the integral defining  $\tilde{Z}_1$  in Eq. (37). We know that this integral becomes simple in the limit that  $C$  is either large or small, so let us start by trying to generate an approximation that will be valid at large  $C$ .

At large  $C$ , the concentration of input molecules can become large, so we expect that the “output noise,”  $\sigma^2 \propto \bar{g}$ , will be dominant. This suggests that we write

$$\begin{aligned} \tilde{Z}_1 &\equiv \int_0^C dx \left\{ \frac{[d\bar{g}(x)/dx]^2}{\bar{g}(x) + x[d\bar{g}(x)/dx]^2} \right\}^{1/2} \\ &\approx \int_0^C dx \left| \frac{d\bar{g}(x)}{dx} \right| \frac{1}{\sqrt{\bar{g}(x)}} \left[ 1 - \frac{1}{2} x \frac{1}{\bar{g}(x)} \left( \frac{d\bar{g}(x)}{dx} \right)^2 + \dots \right]. \end{aligned} \quad (39)$$

To proceed, we note the combination  $dx|d\bar{g}/dx|$ , which invites us to convert this into an integral over  $\bar{g}$ . We use the fact that, for activators described by the Hill function in Eq. (38),

$$x = \frac{K}{c_0} \left( \frac{\bar{g}}{1 - \bar{g}} \right)^{1/n}, \quad (40)$$

$$\frac{d\bar{g}(x)}{dx} = \frac{n}{x}\bar{g}(1-\bar{g}). \quad (41)$$

Substituting, we find

$$\tilde{Z}_1 \approx \int_0^{\bar{g}(C)} \frac{d\bar{g}}{\sqrt{\bar{g}}} \left[ 1 - \frac{c_0 n^2}{2K} \bar{g}^{1-1/n} (1-\bar{g})^{2+1/n} + \dots \right] \quad (42)$$

$$= 2\sqrt{\bar{g}(C)} - \frac{c_0 n^2}{2K} \int_0^{\bar{g}(C)} d\bar{g} \bar{g}^{1/2-1/n} (1-\bar{g})^{2+1/n} + \dots. \quad (43)$$

Again, we are interested in large  $C$ , so we can approximate  $\bar{g}(C) \approx 1 - (K/c_{\max})^n$ . Similarly, the second term in Eq. (43) can be approximated by letting the upper limit on the integral approach 1; the difference between  $\bar{g}(C)$  and 1 generates higher order terms in powers of  $1/C$ . Thus, we have

$$\tilde{Z}_1^{\text{act}} \approx 2 - \left( \frac{K}{c_{\max}} \right)^n - A(n) \frac{c_0 n^2}{2K} + \dots, \quad (44)$$

$$A(n) = \int_0^1 dz z^{1/2-1/n} (1-z)^{2+1/n} \quad (45)$$

$$= \frac{\Gamma(3/2 - 1/n)\Gamma(3 + 1/n)}{\Gamma(9/2)}. \quad (46)$$

The approximate expression for  $\tilde{Z}_1$  expresses the basic compromise involved in optimizing information transmission. On the one hand, we would like  $K$  to be small so that the output runs through its full dynamic range; correspondingly, we want to decrease the term  $(K/c_{\max})^n$ . On the other hand, we want to move the most sensitive part of the input/output relation to higher concentrations, so that we are less sensitive to the input noise; this corresponds to decreasing the term  $\propto c_0/K$ . The optimal compromise is reached at

$$K_{\text{opt}}^{\text{act}} \approx c_{\max} \left[ \frac{nA(n)c_0}{2c_{\max}} \right]^{1/(n+1)}. \quad (47)$$

Parallel arguments yield, for repressors,

$$\tilde{Z}_1^{\text{rep}} \approx 2 - 2 \left( \frac{K}{c_{\max}} \right)^n - B(n) \frac{c_0 n^2}{2K} + \dots, \quad (48)$$

$$K_{\text{opt}}^{\text{rep}} \approx c_{\max} \left[ \frac{nB(n)c_0}{2c_{\max}} \right]^{2/(n+2)}, \quad (49)$$

$$B(n) = \int_0^1 dz z^{1/2+1/n} (1-z)^{2-1/n} \quad (50)$$

$$= \frac{\Gamma(3/2 + 1/n)\Gamma(3 - 1/n)}{\Gamma(9/2)}. \quad (51)$$

The first thing we notice about our approximate results is that the optimal values of  $K$  are almost proportional to  $c_{\max}$ , as one might expect, but not quite—the growth of  $K$  with  $c_{\max}$  is slightly sublinear. Also, one might have expected that  $K$  would be chosen to divide the available dynamic range

into roughly equal “on” and “off” regions, which should maximize the entropy of the output and hence increase the capacity; to achieve this requires  $K_{\text{opt}}/c_{\max} \approx 1/2$ . In fact we see that the ratio  $K_{\text{opt}}/c_{\max}$  is determined by a combination of terms and depends in an essential way on the scale of the input noise  $c_0$ , even though we assume that the maximal concentration is large compared with this scale.

The basic compromise between extending the dynamic range of the outputs and avoiding low input concentrations works differently for activators and repressors. As a result, the optimal values of  $K$  are different in the two cases. From Eq. (37), it is clear that the symmetry between the two types of regulation is broken by the noise term proportional to  $\bar{g}$ . Unless the optimal Hill coefficient for repressors were very much smaller than for activators (and it is not), Eqs. (47) and (49) predict that  $K_{\text{opt}}^{\text{rep}}$  will be smaller than  $K_{\text{opt}}^{\text{act}}$ , in agreement with the numerical results in Fig. 3.

To test these analytic approximations, we can compare the predicted values of  $K_{\text{opt}}$  with those found numerically. There is a slight subtlety, since our analytic results for  $K_{\text{opt}}$  depend on the Hill coefficient  $n$ . We can take this coefficient as known from the numerical optimization or we can use the approximations to  $\tilde{Z}_1$  [as in Eq. (44)] to simultaneously optimize for  $K$  and  $n$ . In contrast to the optimization of  $K$ , however, there is no simple formula for  $n_{\text{opt}}$ , even in our approximation at large  $c_{\max}$ .

Results for the approximate vs numerically exact optimal  $K$  are shown in Fig. 4. As it should, the approximation approaches the exact answer as  $c_{\max}$  becomes large. In fact, the approximation is quite good even at  $c_{\max}/c_0 \sim 10$ , and for activators the error in  $K_{\text{opt}}$  is only  $\sim 15\%$  at  $c_{\max}/c_0 \sim 3$ . Across the full range of  $c_{\max}/c_0 > 1$ , the analytic approximation captures the basic trends:  $K_{\text{opt}}/c_{\max}$  is a slowly decreasing function of  $c_{\max}/c_0$ ,  $K_{\text{opt}}^{\text{act}}$  is larger than  $K_{\text{opt}}^{\text{rep}}$  by roughly a factor of 2, and for both activators and repressors we have  $K_{\text{opt}}$  noticeably smaller than  $c_{\max}/2$ . Similarly good results are obtained for the approximate predictions of the optimal Hill coefficient  $n$  as shown in Fig. 4(b).

As noted above, the large  $c_{\max}$  approximation makes clear that optimizing information transmission is a compromise between using the full dynamic range of outputs and avoiding expression levels associated with large noise at low concentration of the input. The constraint of using the full dynamic range pushes the optimal  $K$  downward; this constraint is stronger for repressors [compare the second terms of Eqs. (44) and (48)] causing the optimal  $K$ s of repressors to be smaller than those of the activators. On the other hand, avoiding input noise pushes the most sensitive part of the expression profile toward high concentrations favoring large  $K$ . The fact that this approximation captures the basic structure of the numerical solution to the optimization problem encourages us to think that this intuitive compromise is the essence of the problem. It is also worth noting that as  $c_{\max}$  increases, activators increase their output range, hence gaining capacity. On the other hand, the output of the repressed systems is small for large  $c_{\max}$  and the output noise thus is large, limiting the increase in capacity compared to the activated genes, as is seen in Fig. 3(c).

In the case of small  $c_{\max}$  it is harder to obtain detailed expressions for  $K$ ; however, we can still gain insight from



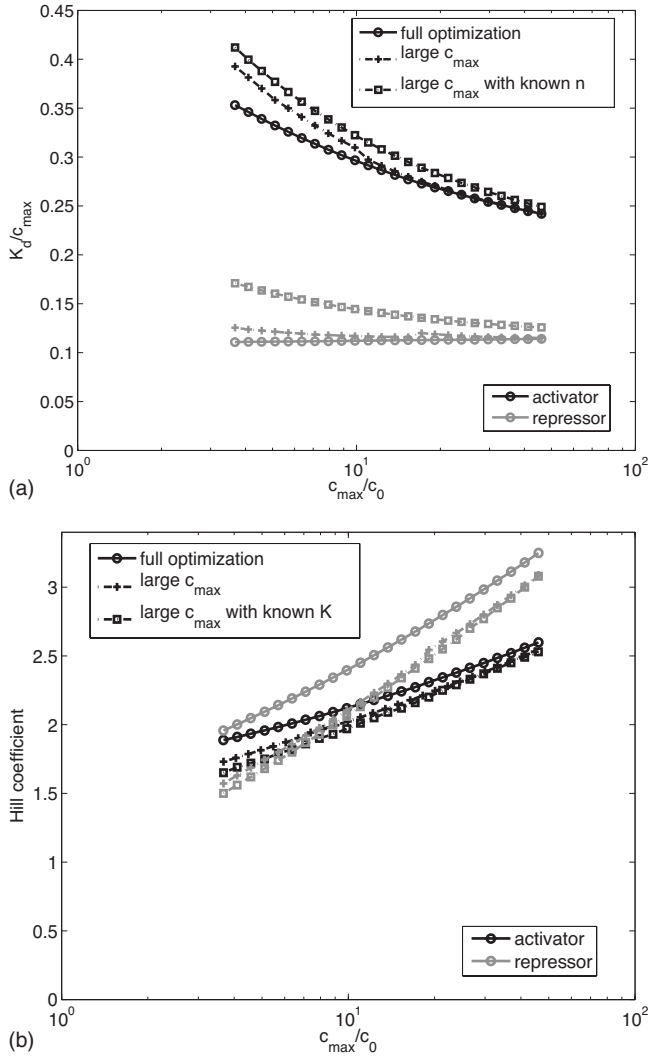


FIG. 4. Approximate results for the optimal values of  $K$  (a) and  $n$  (b) compared with exact numerical results for activators (black lines) and repressors (gray lines). As explained in the text, we can use our analytic approximations to determine, for example, the optimal  $K$  assuming  $n$  is known (large  $c_{\max}$  with known  $n$  results) or we can simultaneously optimize both parameters (large  $c_{\max}$  results); results are shown for both calculations.

the expression for the capacity in this limit. To obtain the large  $c_{\max}$  limit we assumed that  $\bar{g} \gg x(d\bar{g}/dx)^2$  in the denominator of the integrand which defines  $Z_1$ ; to obtain the small  $c_{\max}$  limit we make the opposite assumption:

$$\begin{aligned}
 \tilde{Z}_1 &\approx \int_0^C dx \left\{ \frac{[d\bar{g}(x)/dx]^2}{\bar{g}(x) + x[d\bar{g}(x)/dx]^2} \right\}^{1/2} \\
 &= \int_0^C dx \frac{1}{\sqrt{x}} \left[ \frac{1}{1 + \bar{g}(x)/\{x[d\bar{g}(x)/dx]^2\}} \right]^{1/2} \\
 &\approx \int_0^C dx \frac{1}{\sqrt{x}} \left[ 1 - \frac{x}{2n^2} \frac{1}{g(1-g)^2} + \dots \right], \quad (52)
 \end{aligned}$$

where in the last step we use the relation in Eq. (41). We see that, if  $g$  approaches one, the first correction term will di-

verge. This allows us to predict the essential feature of the optimal solutions at small  $c_{\max}$ , namely, that they do not access the full dynamic range of outputs.

### C. Constraining means

Here we would like to solve the same optimization problem by constraining the mean concentrations rather than imposing a hard constraint on the maximal concentrations; as noted above we can also think of this problem as maximizing information subject to a fixed cost per molecule. To compare results in a meaningful way, we should know how the mean concentration varies as a function of  $c_{\max}$  when we solve the problem with constrained maxima and this is shown in Fig. 5(a). An interesting feature of these results is that mean concentrations are much less than half of the maximal concentration. Also, the mean input concentrations for activator and repressor systems are similar despite different values of the optimal  $K$ . This result shows that for a given dynamic range defined by  $c_{\max}$ , there is an optimal mean input concentration, which is independent of whether the input/output relation is up or down regulating.

Equation (28) shows us how to compute the partition function  $Z_2$  for the case where we constrain the mean concentration of transcription factors and Eq. (30) relates this to the information capacity  $I_2$ . Substituting our explicit expressions for the noise in the case of one input and one output, we have

$$Z_2 = \left[ \frac{N_{\max}}{2\pi e} \right]^{1/2} \tilde{Z}_2, \quad (53)$$

$$\tilde{Z}_2 = \int_0^\infty dc \left\{ \frac{[d\bar{g}(c)/dc]^2}{\bar{g}(c) + cc_0[d\bar{g}(c)/dc]^2} \right\}^{1/2} e^{-ac}. \quad (54)$$

As before, we choose Hill functions for  $\bar{g}(c)$  and maximize  $I_2$  with respect to the parameters  $K$  and  $n$ . This defines a family of optimal solutions parametrized by the Lagrange multiplier  $\alpha$  and we can tune this parameter to match the mean concentration  $\langle c \rangle$ . Using the calibration in Fig. 5(a), we can compare these results with those obtained by optimizing with a fixed maximum concentration. Results are shown in Figs. 5(b)–5(d).

The most important conclusion from Fig. 5 is that constraining mean concentrations and constraining maximal concentrations give—for this simple problem of one input and one output—essentially the same answer. The values of the optimal  $K$ s are almost identical [Fig. 5(c)], as are the actual number of bits that can be transmitted [Fig. 5(b)]. The only systematic difference is in the Hill coefficient  $n$ , where having a fixed maximal concentration drives the optimization toward slightly larger values of  $n$  [Fig. 5(d)] so that more of the dynamic range of outputs is accessed before the system runs up against the hard limit at  $c = c_{\max}$ .

It is interesting that the optimal value of  $K$  is more nearly a linear function of  $\langle c \rangle$  than of  $c_{\max}$ , as we see in Fig. 5(c). To

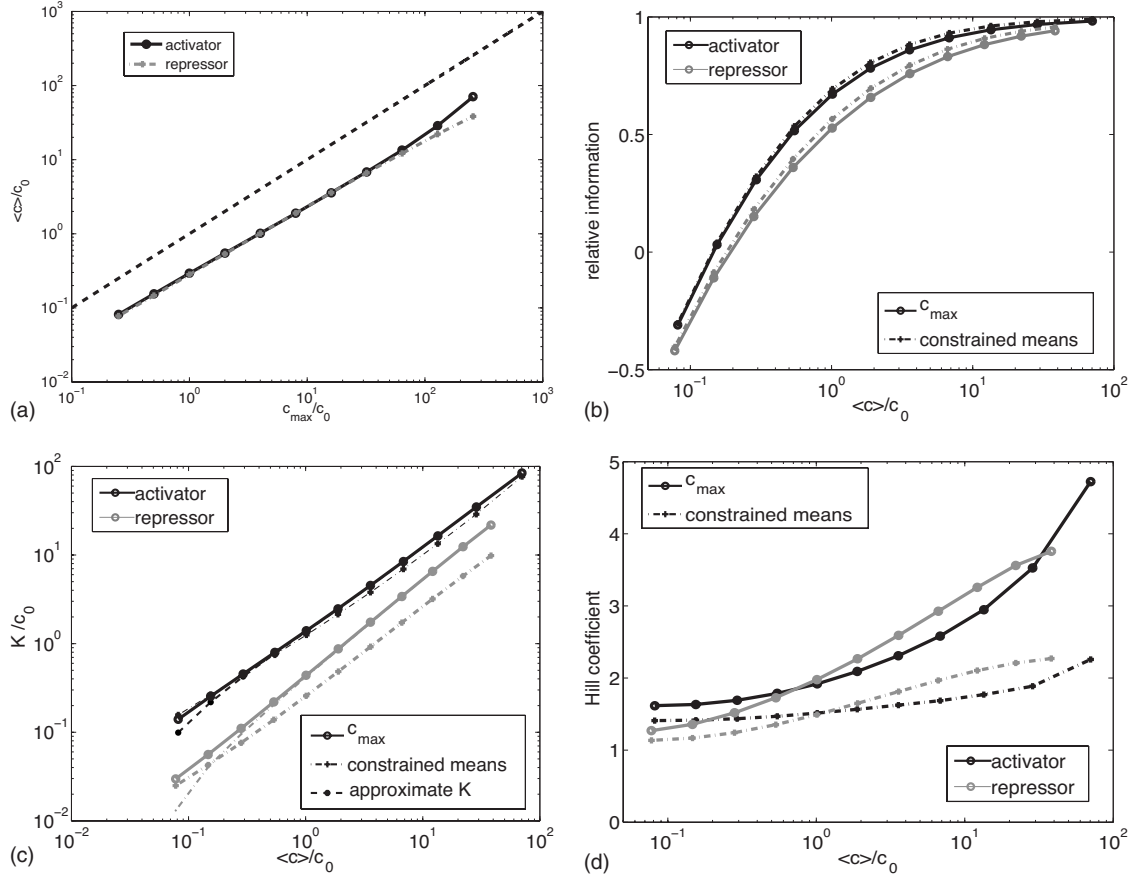


FIG. 5. (a) Mean concentration of the transcription factor when we optimize information transmission subject to a constraint on the maximum concentration. Results are shown for one input and one output, both for activators and repressors. The dashed black line shows equality. (b)–(d) Comparing two formulations of the optimization problem for activators (black lines) and repressors (gray lines) calculated with a finite dynamic range ( $c_{\max}$ —circles and solid lines) and constrained means (crosses and dashed lines). The panels show the relative information in panel (b), the optimal value of  $K$  in panel (c), and the optimal value of the Hill coefficient in panel (d). In panel (c), approximate results for  $K$  are shown as a function of  $\langle c \rangle$  from Eqs. (56) and (58).

understand this, we follow the steps in Sec. III B, expanding the expression for  $\langle c \rangle$  in the same approximation that we used for large  $c_{\max}$ :

$$\begin{aligned}
 \langle c \rangle &= \frac{\int_0^C dc c \left\{ \frac{[d\bar{g}(c)/dc]^2}{\bar{g}(c) + cc_0[d\bar{g}(c)/dc]^2} \right\}^{1/2}}{\int_0^C dc \left\{ \frac{[d\bar{g}(c)/dc]^2}{\bar{g}(c) + cc_0[d\bar{g}(c)/dc]^2} \right\}^{1/2}} \\
 &\approx \frac{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \frac{c}{\sqrt{\bar{g}}} - \frac{1}{2} \int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} n^2 \bar{g}^{-1/2} (1-\bar{g})^2}{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \frac{1}{\sqrt{\bar{g}}} - \frac{1}{2} \int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} n^2 c \bar{g}^{1/2} (1-\bar{g})^2}. \quad (55)
 \end{aligned}$$

$$\begin{aligned}
 \langle c \rangle &= \frac{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/n-1/2} (1-\bar{g})^{-1/n}}{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/2}} \\
 &\times \left[ K + \frac{n^2 \int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/2-1/n} (1-\bar{g})^{2+1/n}}{\int_{\bar{g}(0)}^{\bar{g}(C)} \bar{g}^{-1/2}} + \dots \right] \\
 &= \frac{n^2 \int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/2} (1-\bar{g})^2}{2 \int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/2}}. \quad (56)
 \end{aligned}$$

In the case of an activator,  $c = K/c_0[\bar{g}/(1-\bar{g})]^{1/n}$ , and the leading terms become

To get some intuition for the numerical values of these terms we will assume the integral covers the whole expression range  $\bar{g} \in [0, 1]$ , and  $n=3$ . Then this expression simplifies to

$$\langle c \rangle \approx 0.86K + 0.52, \quad (57)$$

so we understand how this simple result emerges, at least asymptotically at large  $c_{\max}$ .

In the case of repressors the leading terms are

$$\begin{aligned} \langle c \rangle = & \frac{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/n-1/2} (1-\bar{g})^{1/n}}{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/2}} \\ & \times \left[ K + \frac{n^2}{2} \frac{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{1/2+1/n} (1-\bar{g})^{2-1/n}}{\int_{\bar{g}(0)}^{\bar{g}(C)} \bar{g}^{-1/2}} + \dots \right] \\ & - \frac{n^2}{2} \frac{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{1/2} (1-\bar{g})^2}{\int_{\bar{g}(0)}^{\bar{g}(C)} d\bar{g} \bar{g}^{-1/2}}. \end{aligned} \quad (58)$$

As in the case of the activator, making the rough approximation that  $n=3$  and  $\bar{g} \in [0, 1]$  allows us to get some intuition for this large  $c_{\max}$  result:

$$\langle c \rangle \approx 2.8K + 1.19. \quad (59)$$

These extremely crude estimates do predict the basic linear trends in Fig. 5(c), including the fact that for a given value of the mean concentration, the repressor has a smaller  $K$  than the activator.

Before leaving this section, we should return to the question of constraining mean outputs, as well as mean inputs. We have measured the input concentration in absolute units (or relative to the physical scale  $c_0$ ), so when we constrain the mean input we really are asking that the system use only a fixed mean number of molecules. In contrast, we have measured outputs in relative units, so that the maximum of  $\bar{g}(c)$  is 1. If we want to constrain the mean number of output molecules, we need to fix not  $\langle g \rangle$ , but rather  $N_{\max} \langle g \rangle$ , since the factor of  $N_{\max}$  brings us back to counting the molecules in absolute terms [60]. Thus, exploring constrained mean output requires us to view  $N_{\max}$  (and hence the scale  $c_0$ ) as an extra adjustable parameter.

By itself, adding  $N_{\max}$  as an additional optimization parameter makes our simple problem more complicated, but does not seem to add much insight. In principle it would allow us to discuss the relative information gain on adding extra input vs output molecules, with the idea that we might find optimal information transmission subject to some net resource constraint; for initial results in this direction see Ref. [41]. In networks with feedback, the target genes also act as transcription factors and these tradeoffs should be more interesting. We will return to this problem in subsequent papers.

#### IV. MULTIPLE OUTPUTS

When the single transcription factor at the input of our model system has multiple independent target genes, and we constrain the maximal concentrations, the general form of the information capacity in the small noise limit is given by Eq. (24),

$$\begin{aligned} Z_1 = & \int_0^{c_{\max}} dc \left\{ \frac{1}{2\pi e} \sum_{i=1}^M \frac{1}{\sigma_i^2(c)} \left[ \frac{d\bar{g}_i(c)}{dc} \right]^2 \right\}^{1/2} \\ = & \left[ \frac{N_{\max}}{2\pi e} \right]^{1/2} \int_0^{c_{\max}} dc \left\{ \sum_{i=1}^M \frac{[d\bar{g}_i(c)/dc]^2}{\bar{g}_i(c) + c c_0 [d\bar{g}_i(c)/dc]^2} \right\}^{1/2}, \end{aligned} \quad (60)$$

where we assume for simplicity that the basic parameters  $N_{\max}$  and  $D\ell\tau$  are the same for all the target genes. Once again,  $c_0 = N_{\max}/D\ell\tau$  provides a natural unit of concentration. We limit ourselves to an extended discussion of the case with a hard upper bound,  $c_{\max}$ , to the dynamic range of the input. As in the case of a single output, the calculation with a constrained mean input concentration gives essentially the same results.

To get an initial feeling for the structure of the problem, we try the case of five target genes, all of which are activated by the transcription factor. Then,

$$\bar{g}_i(c) = \frac{c^{n_i}}{c^{n_i} + K_i^{n_i}}, \quad (61)$$

and we can search numerically for the optimal settings of all the parameters  $\{K_i, n_i\}$ . Results are shown in Fig. 6. A striking feature of the problem is that, for small values of the maximal concentration  $C = c_{\max}/c_0$ , the optimal solution is actually to have all five target genes be completely redundant, with identical values of  $K_i$  and  $n_i$ . As  $c_{\max}$  increases, this redundancy is lifted and the optimal solution becomes a sequence of target genes with staggered activation curves, in effect ‘‘tiling’’ the input domain  $0 < c < c_{\max}$ . To interpret these results, we realize that for small maximal concentration the input noise dominates and the optimal strategy for  $M$  genes is to ‘‘replicate’’ one well-placed gene  $M$  times: having  $M$  independent and redundant readouts (with identical  $K$  and  $n$ ) of the input concentration will decrease the noise by a factor of  $\sqrt{M}$ . However, as the dynamic range increases and output noise has a chance to compete with the input noise, more information can be transmitted by using  $M$  genes to probe the input at different concentrations, thereby creating a cascade of genes that get activated at successively higher and higher input levels. The transition between these two readout strategies is described in more detail below.

To look more closely at the structure of the problem, we drop down to consider two target genes. Then there are three possibilities—two activators (AA), two repressors (RR), and one of each (AR). For each of these discrete choices, we have to optimize two exponents  $(n_1, n_2)$  and two half-maximal points  $(K_1, K_2)$ . In Fig. 7 we show how  $\bar{Z}_1$  varies in the  $(K_1, K_2)$  plane assuming that at every point we choose the optimum exponents; the different quadrants correspond to the different discrete choices of activator and repressor. The

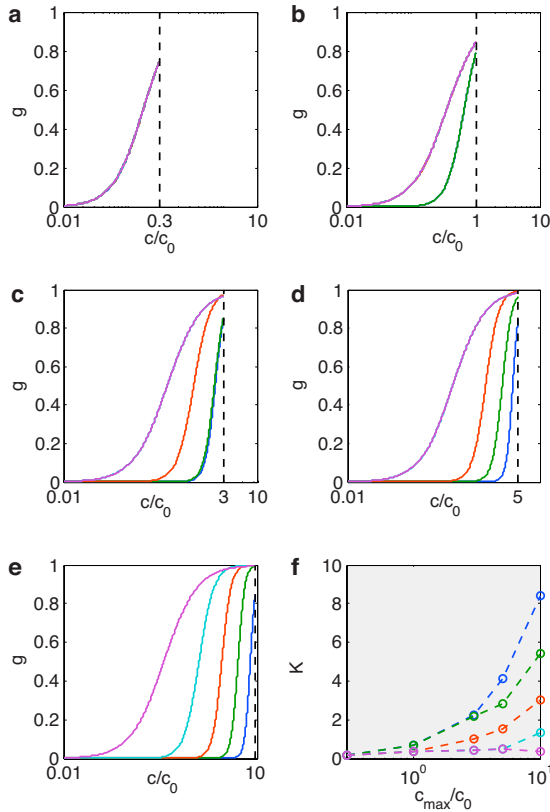


FIG. 6. (Color online) Optimal input/output relations for the case of five independent target genes activated by the TF at concentration  $c$ . Successive panels [(a)–(e)] correspond to different values of the maximal input concentration as indicated ( $C=0.3, 1, 3, 5, 10$ ). Panel (f) summarizes the optimal values of the  $K_i$  as a function of  $C=c_{\max}/c_0$ : as  $C$  is increased, the  $K_i$  of the fully redundant input/output relations for  $C=0.3$  bifurcate such that at  $C=10$  the genes tile the whole input range.

results show clearly how the redundant ( $K_1=K_2$ ) solutions at low values of  $c_{\max}$  bifurcate into asymmetric ( $K_1 \neq K_2$ ) solutions at larger values of  $c_{\max}$ ; the critical value of  $c_{\max}$  is different for activators and repressors. This bifurcation structure is summarized in Fig. 8, where we also see that, for each value of  $c_{\max}$ , the three different kinds of solutions (AA, RR, and AR) achieve information capacities that differ by less than 0.1 bits.

The information capacity is an integral of the square root of a sum of terms, one for each target gene [Eq. (60)]. Thus, if we add redundant copies of a single gene, all with the same values of  $K$  and  $n$ , the integral  $Z_1$  will scale as  $\sqrt{M}$ , where  $M$  is the number of genes. In particular, as we go from one to two target genes,  $Z$  would increase by a factor  $\sqrt{2}$  and hence the information capacity,  $\log_2 Z$ , would increase by one half bit; more generally, with  $M$  redundant copies, we have  $(1/2)\log_2 M$  bits of extra information relative to having just one gene. On the other hand, if we could arrange for two target genes to make nonoverlapping contributions to the integral, then two genes could have a value of  $Z$  that is twice as large as for one gene, generating an extra bit rather than an extra half bit. In fact a full factor of 2 increase in  $Z$  is not achievable because once the two target genes are sampling different regions of concentration they are making different

tradeoffs between the input and output noise terms; since the one gene had optimized this tradeoff, bifurcating into two distinguishable targets necessarily reduces the contribution from each target. Indeed, if the maximal concentration is too low then there is no “space” along the  $c$  axis to fit two distinct activation (or repression) curves, and this is why low values of  $c_{\max}$  favor the redundant solutions.

Figure 9(a) shows explicitly that when we increase the number of target genes at low values of  $c_{\max}$ , the optimal solution is to use the genes redundantly and hence the gain in information is  $(1/2)\log_2 M$ . At larger values of  $c_{\max}$ , going from one target to two targets one can gain more than half a bit, but this gain is bounded by 1 bit, and indeed over the range of  $c_{\max}$  that we explore here the full bit is never quite reached.

We can take a different slice through the parameter space of the problem by holding the number of target genes fixed and varying  $c_{\max}$ . With a single target gene, we have seen (Fig. 3) that the information capacity saturates rapidly as  $c_{\max}$  is increased above  $c_0$ . We might expect that, with multiple target genes, it is possible to make better use of the increased dynamic range and this is what we see in Fig. 9(b).

For a system with many target genes, it is illustrative to plot the optimal distribution of input levels,  $P_{TF}^*(c) \propto \sigma_c^{-1}(c)$ . Figure 10 shows the results for the case of  $M=2, 3, \dots, 9$  genes at low ( $C=0.3$ ) and high ( $C=30$ ) input dynamic range. At low input dynamic range the distributions for various  $M$  collapse onto each other (because the genes are redundant), while at high  $C$  increasing the number of genes drives the optimal distribution closer to  $\propto c^{-1/2}$ . We recall that the input noise is  $\sigma_c \propto \sqrt{c}$ , so this shows that, as the number of targets becomes large, the input noise becomes dominant over a wider and wider dynamic range.

Finally, one can ask how finely tuned the input/output relations for the particular genes need to be in a maximally informative system. To consider how the capacity of the system changes when the parameters of the input/output relations change slightly, we analyzed the (Hessian) matrix of second derivatives of the information with respect to fractional changes in the various parameters; we also made more explicit maps of the variations in information with respect to the individual parameters and sampled the variations in information that result from random variations in the parameters within some range. Results for a two gene system are illustrated in Fig. 11.

The first point concerns the scale of the variations—20% changes in parameters away from the optimum result in only  $\sim 0.01$  bits of information loss, and this is true both at low  $c_{\max}$  where the solutions are redundant and at high  $c_{\max}$  where they are not. Interestingly, the eigenmodes of the Hessian reveal that in the asymmetric case the capacity is most sensitive to variations in the larger  $K$ . The second most sensitive (much weaker than the first) direction is a linear combination of both of the parameters  $K$  and  $n$  for the gene which is activated at lower concentrations. Perhaps surprisingly, this means that genes which activate at higher  $K$  need to have their input/output relations positioned with greater accuracy along the  $c$  axis, even in fractional terms. If we think of  $K \sim e^{-F/k_B T}$ , where  $F$  is the binding (free) energy of the transcription factor to its specific target site along the

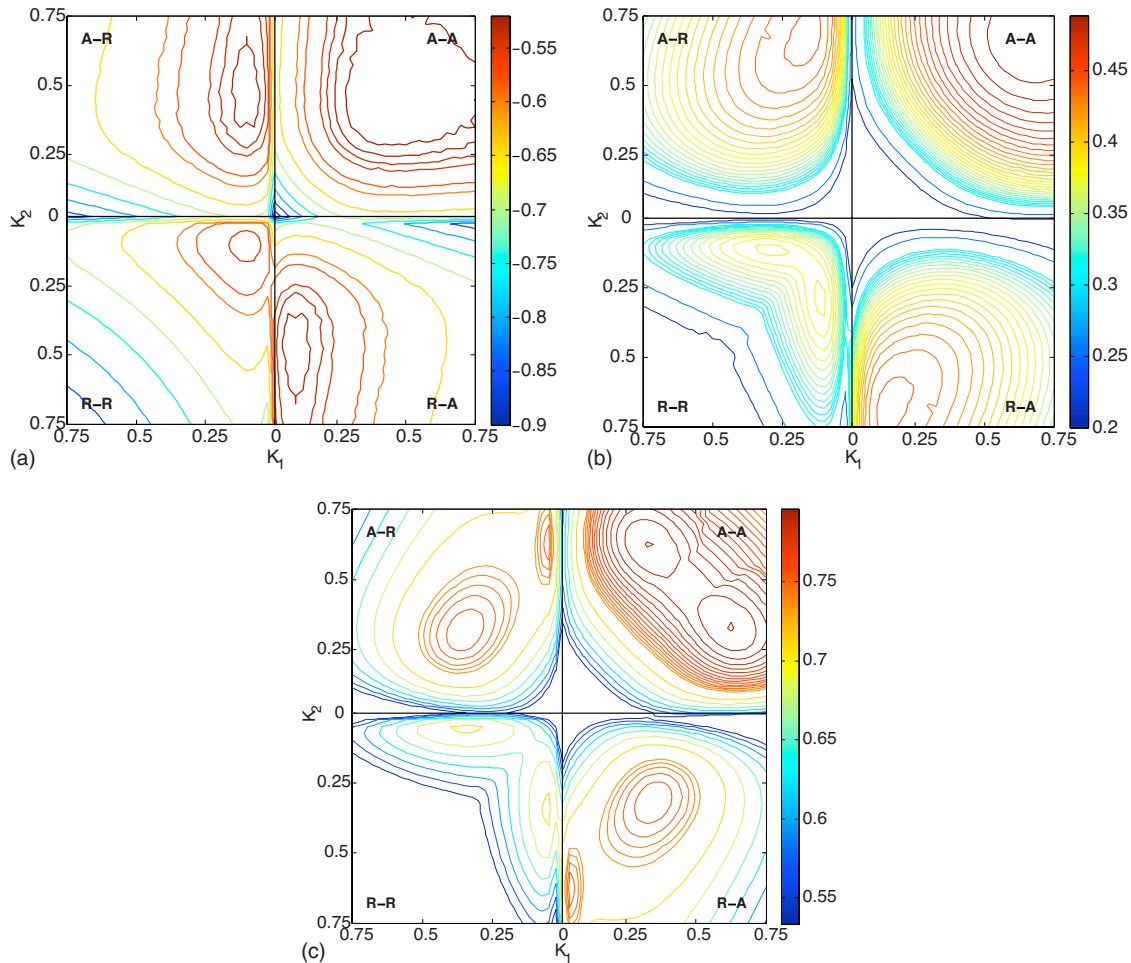


FIG. 7. (Color online) The case of two target genes. The maps show contour plots of relative information ( $\log_2 \tilde{Z}_1$ ) as a function of the  $K$  values of the two genes:  $K_1$  and  $K_2$ . In each map, the upper right quadrant (A-A) contains solutions where both genes are activated by a common TF, in the lower left quadrant (R-R) both genes are repressed, and the other two quadrants (A-R) contain an activator-repressor mix. The maximal concentration of the input molecules is fixed at  $c_{\max}/c_0=0.1$  in map (a), at 0.5 in map (b), and at 1 in map (c). We see that, for example, only at the highest value of  $c_{\max}$  does the two-activator solution in the upper right quadrant correspond to distinct values of  $K_1$  and  $K_2$ ; at lower values of  $c_{\max}$  the optimum is along the “redundant” line  $K_1=K_2$ . The redundancy is lifted at lower values of  $c_{\max}$  in the case of repressors, as we see in the lower left quadrants, and the mixed activator/repressor solutions are always asymmetric. At large  $c_{\max}$  we also see that there are two distinct mixed solutions.

genome, another way of stating this result is that weaker binding energies (smaller  $F$ ) must be specified with greater precision to achieve a criterion level of performance. Finally, if we allow parameters to vary at random, we see [Figs. 11(c) and 11(d)] that the effects on information capacity are extremely small as long as these variations are bounded, so that the range of the natural log of the parameters is significantly less than 1. If we allow larger fluctuations, there is a transition to a much broader distribution of information capacities, with a substantial loss relative to the optimum.

## V. DISCUSSION

The ability of cells to control the expression levels of their genes is central to growth, development, and survival. In this work we have explored perhaps the simplest model for this control process, in which changes in the concentration of a single transcription factor protein modulate the expression of

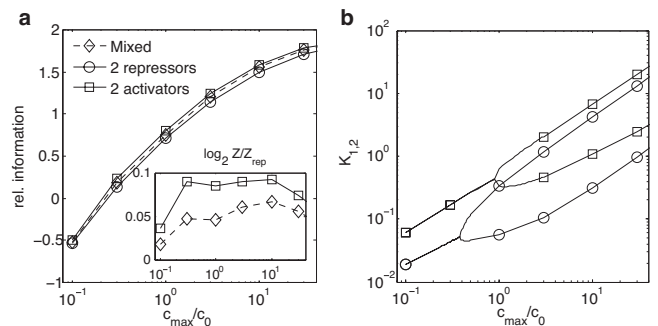


FIG. 8. The relative information for stable solutions for two genes as a function of  $c_{\max}$  [panel (a)]. The inset shows the difference in information transmission for two activators and the mixed case relative to the two repressors. In panel (b), the optimal  $K_1$  and  $K_2$  are plotted as a function of  $c_{\max}$  for two activators (squares) and two repressors (circles). The bifurcation in  $K$  is a continuous transition that happens at lower  $c_{\max}$  in the case of two repressors.

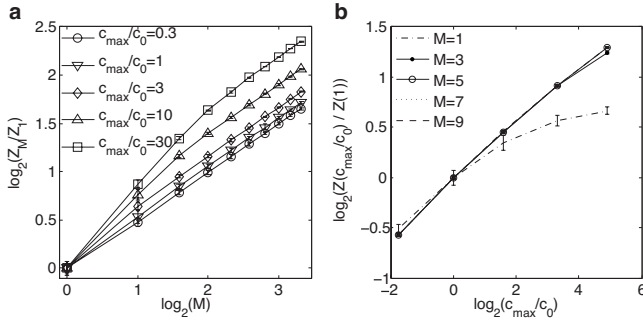


FIG. 9. The relative information for different values of  $c_{\max}$  as a function of the number of genes,  $M$ , shown in panel (a). At low  $c_{\max}$  the genes are redundant and so the capacity grows as  $(1/2)\log_2 M$ ; at high  $c_{\max}$ , the increase in capacity is larger but bounded from above by 1 bit. The differences in information for various combinations of activators and repressors are comparable to the size of the plot symbols. In panel (b), the relative information for different numbers of genes as a function of  $c_{\max}$ . At higher  $M$ , the system can make better use of the input dynamic range.

one or more genes by binding to specific sites along the DNA. Such models have many parameters, notably the binding energies of the transcription factor to the different target sites and the interactions or cooperativity among factors bound to nearby sites that contribute to the control of the same gene. This rapid descent from relatively simple physical pictures into highly parametrized models is common to

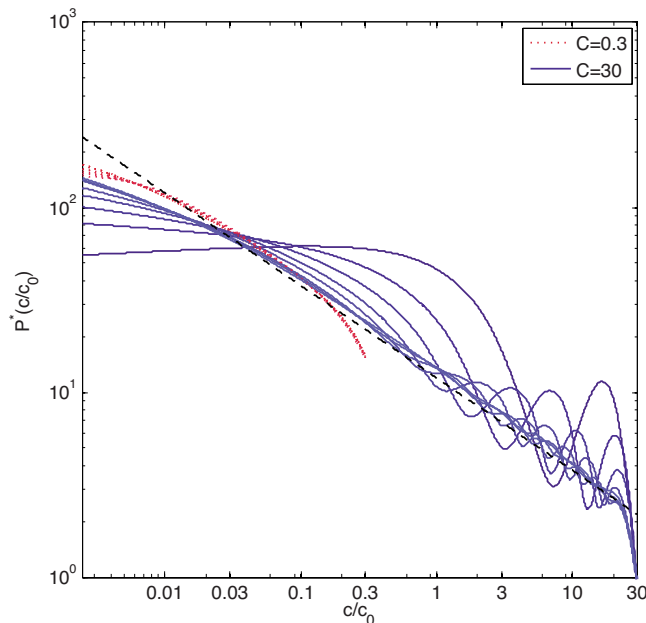


FIG. 10. (Color online) The optimal probability distribution of inputs,  $P_{TF}^*(c)$ . In red (dotted line), plotted for  $C=0.3$ . In blue (solid line), plotted for  $C=30$ . Different lines correspond to solutions with 2, 3, ..., 9 genes. At low  $C$  (red dotted line), the genes are degenerate and the input distribution is independent of the number of genes. At high  $C$  (blue solid line), where the genes tile the concentration range, the optimal input distribution approaches  $(c/c_0)^{-1/2}$  (dashed line) as the number of target genes increases.

most modern attempts at quantitative analysis of biological systems. Our goal in this work is to understand whether these many parameters can be determined by appeal to theoretical principles rather than solely by fitting to data.

We begin our discussion with a caveat. Evidently, deriving the many parameters that describe a complex biological system is an ambitious goal and what we present here is at best a first step. By confining ourselves to systems in which one transcription factor modulates the expression of many genes, with no further inputs or interactions, we almost certainly exclude the possibility of direct, quantitative comparisons with real genetic control networks. Understanding this simpler problem, however, is a prerequisite to analysis of more complex systems, and, as we argue here, sufficient to test the plausibility of our theoretical approach.

The theoretical principle to which we appeal is the optimization of information transmission. In the context of genetic control systems, we can think of information transmission as a measure of control power—if the system can transmit  $I$  bits of information, then adjustment of the inputs allows the cell to access, reliably,  $2^I$  distinguishable states of gene expression. In unicellular organisms, for example, these different states could be used to match cellular metabolism to the available nutrients, while in the developing embryo of a multicellular organism these different states could be the triggers for emergence of different cell types or spatial structures; in either case, it is clear that information transmission quantifies our intuition about the control power or (colloquially) complexity that the system can achieve. Although one could imagine different measures, specialized to different situations, we know from Shannon that the mutual information is the unique measure that satisfies certain plausible conditions and works in all situations [30,31].

Information transmission is limited by noise. In the context of genetic control systems, noise is significant because the number of molecules involved in the control process is small and basic physical principles dictate the random behavior of the individual molecules. In this sense, the maximization of information transmission really is the principle that organisms should extract maximum control power from a limited number of molecules. Analysis of experiments on real control elements suggests that the actual number of molecules used by these systems sets a limit of 1–3 bits on the capacity of a transcription factor to control the expression level of one gene, that significant increases in this capacity would require enormous increases in the number of molecules, and that, at least in one case, the system can achieve  $\sim 90\%$  of its capacity [41,42]. Although these observations are limited in scope, they suggest that cells may need to operate close to the informational limits set by the number of molecules that they can afford to devote to these genetic control processes.

The strategy needed to optimize information transmission depends on the structure of the noise in the system. In the case of transcriptional control, there are two irreducible noise sources, the random arrival of transcription factors at their target sites and the shot noise in the synthesis and degradation of the output molecules (mRNA or protein). The interplay between these noise sources sets a characteristic scale for the concentration of transcription factors,

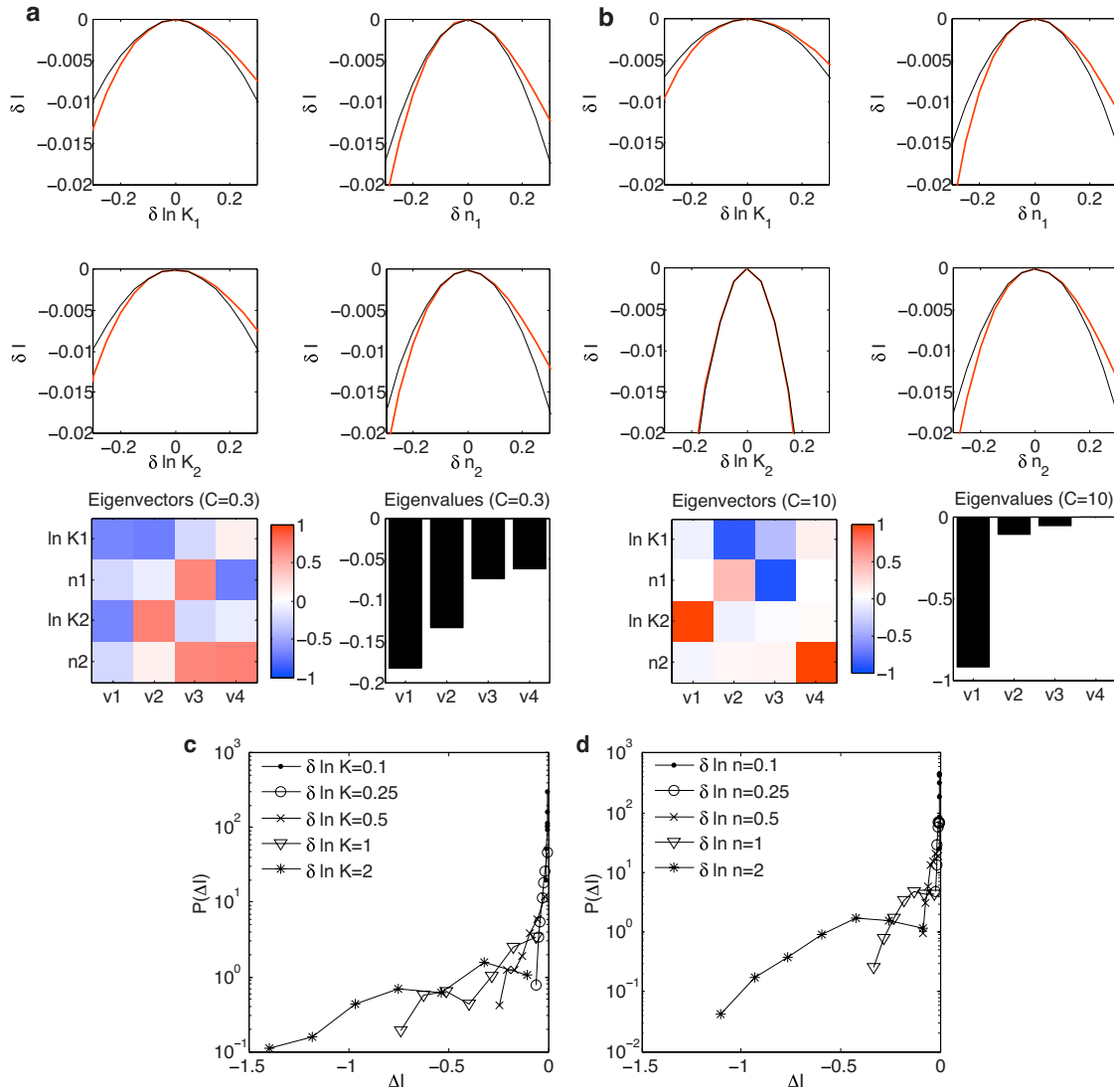


FIG. 11. (Color online) Parameter variations away from the optimum. Results are shown for a two gene system focusing on the solution with two activators. (a) Analysis of the Hessian matrix for  $c_{\max}/c_0=0.3$ , where the two genes are redundant. Top four panels show the variation in information ( $\delta I$  in bits) along each dimension of the parameter space (thick red line) and the quadratic approximation. (b) As in (a), but with  $c_{\max}/c_0=10$ , where the optimal solution is nonredundant. We also show the eigenvectors and eigenvalues of the Hessian matrix. (c) Distribution of information loss  $\Delta I$  when the parameters  $K_1$  and  $K_2$  are chosen at random from a uniform distribution in  $\ln K$ , with widths as shown; here  $c_{\max}/c_0=10$ . (d) As in (c), but for variations in the Hill coefficients  $n_1$  and  $n_2$ .

$c_0 \sim 15\text{--}150$  nM. If the maximum available concentration is too much larger or smaller than this scale, then the optimization of information transmission becomes degenerate and we lose predictive power. Further,  $c_0$  sets the scale for diminishing returns, such that increases in concentration far beyond this scale contribute progressively smaller amounts of added information capacity. Thus, with any reasonable cost for producing the transcription factor proteins, the optimal tradeoff between bits and cost will set the mean or maximal concentration of transcription factors in the range of  $c_0$ . Although only a very rough prediction, it follows without detailed calculation, and it is correct (Table I).

The optimization of information transmission is largely a competition between the desire to use the full dynamic range of outputs and the preference for outputs that can be generated reproducibly, that is, at low noise. Because of the com-

petition of noise sources, this competition has nontrivial consequences, even for a single transcription factor controlling one gene. As we consider the control of multiple genes, the structure of the solutions becomes richer. Activators and repressors are both possible and can achieve nearly identical information capacities. With multiple target genes, all the combinations of activators and repressors also are possible [61]. This suggests that, generically, there will be exponentially many networks that are local optima, with nearly identical capacities, making it possible for a theory based on optimization to generate diversity.

For a limited range of input transcription factor concentrations, the solutions which optimize information transmission involve multiple redundant target genes. Absent this result, the observation of redundant targets in real systems would be interpreted as an obvious sign of nonoptimality, a

remnant of evolutionary history, or perhaps insurance against some rare catastrophic failure of one component. As the available range of transcription factor concentrations becomes larger, optimal solutions diversify, with the responses of the multiple target genes tiling the dynamic range of inputs. In these tiling solutions, targets that require higher concentrations to be activated or repressed also are predicted to exhibit greater cooperativity; in such an optimized system one thus should find some genes controlled by a small number of strong binding sites for the transcription factor and other genes with a large number of weaker sites.

To a large extent, the basic structure of the (numerically) optimal solutions can be recovered analytically through various approximation schemes. These analytic approximations make clear that the optimization really is driven by a conflict between using the full dynamic range of outputs and avoiding states with high intrinsic noise. In particular, this means that simple intuitions based on maximizing the entropy of output states, which are correct when the noise is unstructured [34], fail. Thus, almost all solutions have the property that at least one target gene is not driven through the full dynamic range of its outputs, and even with one gene the midpoint of the optimal activation curve can be far from the midpoint of the available range of inputs. The interplay between different noise sources also breaks the symmetry between activators and repressors, so that repressors optimize their information transmission by using only a small fraction of the available input range.

The predictive power of our approach depends on the existence of well defined optima. At the same time, it would be difficult to imagine evolution tuning the parameters of these models with extreme precision, so the optima should not be too sharply defined. Indeed, we find that optima are clear but broad. In the case of multiple genes, random  $\sim 25\%$  variations in parameters around their optima result in only tiny fractions of a bit of information loss, but once fluctuations become larger than this the information drops precipitously. Looking more closely, we find that proper placement of the activation curves at the upper end of the input range is more critical implying that it is actually the weaker binding sites whose energies need to be adjusted more carefully (perhaps contrary to intuition).

With modest numbers of genes, the optimization approach we propose here has the promise of making rather detailed predictions about structure of the input/output relations, generating what we might think of as a spectrum of  $K$ s and  $n$ s. In the limit of larger networks, we might expect this spectrum to have some universal properties and we see hints of this in Fig. 10. Here, as we add more and more target genes, the optimal distribution of inputs approaches an asymptote  $P_{TF}(c) \propto 1/\sqrt{c}$ ; more of this limiting behavior is accessible if the available dynamic range of inputs is larger. This is the form we expect if the effective noise is dominated by the input noise,  $\sigma_c \propto \sqrt{c}$ . Thus, adding more targets and placing them optimally allows the system to suppress output noise and approach ever more closely the fundamental limits set by the physics of diffusion.

Although there are not so many direct physical measurements specifying the input/output relations of genetic regulatory elements, there are many systems in which there is

evidence for tiling of the concentration axis by a set of target genes, all regulated by the same transcription factor, along the lines predicted here [62]. For example, in quorum sensing by bacteria, the concentrations of extracellular signaling molecules are translated internally into different concentrations of LuxR, which acts as a transcription factor on a number of genes, and these can be classified as being responsive to low, intermediate, and high levels of LuxR [63]. Similarly, the decision of *Bacillus subtilis* to sporulate is controlled by the phosphorylated form of the transcription factor Spo0A, which regulates the expression of  $\sim 30$  genes as well as an additional 24 multigene operons [64]. For many of these targets the effects of SpoA  $\sim$  P are direct and the sensitivity to high vs low concentrations can be correlated with the binding energies of the transcription factor to the particular promoters [65]. In yeast, the transcription factor Pho4 is a key regulator of phosphate metabolism and activates targets such as *pho5* and *pho84* at different concentrations [66]. All of these are potential test cases for the theoretical approach we have outlined here (each with its own complications), but a substantially new level of quantitative experimental work would be required to test the theory meaningfully.

The classic example of multiple thresholds in the activation of genes by a single transcription factor is in embryonic development [3,4]. In this context, spatial gradients in the concentration of transcription factors and other signaling molecules mean that otherwise identical cells in the same embryo experience different inputs. If multiple genes are activated by the same transcription factor but at different thresholds, then smooth spatial gradients can be transformed into sharper “expression domains” that provide the scaffolding for more complex spatial patterns. Although controversies remain about the detailed structure of the regulatory network, the control of the “gap genes” in the *Drosophila* embryo by the transcription factor Bicoid seems to provide a clear example of these ideas [4,67–71]. Recent experimental work [16,17] suggests that it will be possible to make absolute measurements of (at least) Bicoid concentrations, and to map the input/output relations and noise in this system, holding out the hope for more quantitative comparison with theory.

Finally, we look ahead to the more general problem in which multiple target genes are allowed to interact. Absent these interactions, even our optimal solutions have a strong degree of redundancy—as the different targets turn on at successively higher concentrations of the input, there is a positive correlation and hence redundancy among the signals that they convey. This redundancy could be removed by mutually repressive interactions among the target genes, increasing the efficiency of information transmission in much the same way that lateral inhibition or center-surround organization enhances the efficiency of neural coding in the visual system [33,35]. It is known that such mutually repressive interactions exist, for example, among the gap genes in the *Drosophila* embryo [72]. The theoretical challenge is to see if these observed structures can be derived, quantitatively, from the optimization of information transmission.



## ACKNOWLEDGMENTS

We thank T. Gregor, J. B. Kinney, P. Mehta, T. Mora, S. F. Nørrelykke, E. D. Siggia, and especially C. G. Callan for helpful discussions. Work at Princeton was supported in part by NSF Grant No. PHY-0650617, and by NIH Grants No.

P50 GM071508 and No. R01 GM077599. G.T. was supported in part by NSF Grants No. DMR04-25780 and No. IBN-0344678 and by the Vice Provost for Research at the University of Pennsylvania. W.B. also thanks his colleagues at the University of Rome, La Sapienza for their hospitality during a portion of this work.

- 
- [1] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, 1997).
- [2] M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Press, New York, 2002).
- [3] L. Wolpert, *J. Theor. Biol.* **25**, 1 (1969).
- [4] P. A. Lawrence, *The Making of a Fly: The Genetics of Animal Design* (Blackwell, Oxford, 1992).
- [5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. D. Swain, *Science* **297**, 1183 (2002).
- [6] E. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, *Nat. Genet.* **31**, 69 (2002).
- [7] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, *Nature (London)* **422**, 633 (2003).
- [8] Y. Setty, A. E. Mayo, M. G. Surette, and U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7702 (2003).
- [9] J. M. Raser and E. K. O'Shea, *Science* **304**, 1811 (2004).
- [10] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, *Science* **307**, 1962 (2005).
- [11] J. M. Pedraza and A. van Oudenaarden, *Science* **307**, 1965 (2005).
- [12] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, *Cell* **123**, 1025 (2005).
- [13] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, L. DeRisi, and J. S. Weissman, *Nature (London)* **441**, 840 (2006).
- [14] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, *Nat. Genet.* **38**, 636 (2006).
- [15] T. Kuhlman, Z. Zhang, M. H. Saier, Jr., and T. Hwa, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6043 (2007).
- [16] T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank, *Cell* **130**, 141 (2007).
- [17] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, *Cell* **130**, 153 (2007).
- [18] A. Arkin, J. Ross, and H. H. McAdams, *Genetics* **149**, 1633 (1998).
- [19] T. Kepler and T. Elston, *Biophys. J.* **81**, 3116 (2001).
- [20] P. S. Swain, M. B. Elowitz, and E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795 (2002).
- [21] W. Bialek and S. Setayeshgar, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10040 (2005).
- [22] J. Paulsson, *Nature (London)* **427**, 415 (2004).
- [23] A. M. Walczak, M. Sasai, and P. G. Wolynes, *Biophys. J.* **88**, 828 (2005).
- [24] N. E. Buchler, U. Gerland, and T. Hwa, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9559 (2005).
- [25] S. Tanase-Nicola, P. B. Warren, and P. R. ten Wolde, *Phys. Rev. Lett.* **97**, 068102 (2006).
- [26] G. Tkačik and W. Bialek, *Phys. Rev. E* **79**, 051901 (2009).
- [27] M. J. Morelli, R. J. Allen, S. Tanase-Nicola, and P. R. ten Wolde, *Biophys. J.* **94**, 3413 (2008).
- [28] W. Bialek and S. Setayeshgar, *Phys. Rev. Lett.* **100**, 258101 (2008); e-print arXiv:q-bio.MN/0601001.
- [29] G. Tkačik, T. Gregor, and W. Bialek, *PLoS ONE* **3**, e2774 (2008); e-print arXiv:q-bio.MN/0701002.
- [30] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948); reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).
- [31] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley, New York, 1991).
- [32] F. Attneave, *Psychol. Rev.* **61**, 183 (1954).
- [33] H. B. Barlow, *Sensory Mechanisms, the Reduction of Redundancy, and Intelligence*, in Proceedings of the Symposium on the Mechanization of Thought Processes Vol. 2, edited by D. V. Blake and A. M. Uttley (HM Stationery Office, London, 1959), pp. 537–574.
- [34] S. B. Laughlin, *Z. Naturforsch. C* **36c**, 910 (1981).
- [35] J. J. Atick and A. N. Redlich, *Neural Comput.* **2**, 308 (1990).
- [36] S. Smirnakis, M. J. Berry II, D. K. Warland, W. Bialek, and M. Meister, *Nature (London)* **386**, 69 (1997).
- [37] N. Brenner, W. Bialek, and R. de Ruyter van Steveninck, *Neuron* **26**, 695 (2000).
- [38] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck, *Nature (London)* **412**, 787 (2001).
- [39] M. Maravall, R. S. Petersen, A. L. Fairhall, E. Arabzadeh, and M. E. Diamond, *PLoS Biol.* **5**, e19 (2007).
- [40] B. Wark, B. N. Lundstrom, and A. Fairhall, *Curr. Opin. Neurobiol.* **17**, 423 (2007).
- [41] G. Tkačik, C. G. Callan, Jr., and W. Bialek, *Phys. Rev. E* **78**, 011910 (2008).
- [42] G. Tkačik, C. G. Callan, Jr., and W. Bialek, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12265 (2008).
- [43] E. Ziv, I. Nemenman, and C. H. Wiggins, *PLoS One* **2**, e1077 (2007); e-print arXiv:q-bio/0612041.
- [44] E. Emberly, *Phys. Rev. E* **77**, 041903 (2008).
- [45] F. Tostevin and P. R. ten Wolde, *Phys. Rev. Lett.* **102**, 218101 (2009).
- [46] Strictly speaking, Shannon considered the question “how much does the output tell us about the input?” which is almost the same as the question about control power posed here.
- [47] In general, if we allow for interactions,  $P(\{g_i\}|c) = \frac{1}{(2\pi)^{M/2}} \exp\{\frac{1}{2} \ln \det(\mathcal{K}) - \frac{1}{2} \sum_{i,j=1}^M [g_i - \bar{g}_i(c)] \mathcal{K}_{ij} [g_j - \bar{g}_j(c)]\}$ , where  $\mathcal{K}$  measures the (inverse) covariance of the fluctuations or noise in the expression levels at fixed input,  $\langle [g_i - \bar{g}_i(c)] [g_j - \bar{g}_j(c)] \rangle = (\mathcal{K}^{-1})_{ij}$ . In the present paper we restrict the discussion to independent genes  $\mathcal{K}_{ij} = \delta_{ij} / \sigma_i^2$ .
- [48] We are being a bit sloppy here, since the input  $c$  is a continuous variable, but the intuition is correct.

- [49] L. Bintu, N. E. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, *Curr. Opin. Genet. Dev.* **15**, 116 (2005).
- [50] L. Bintu, N. E. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, *Curr. Opin. Genet. Dev.* **15**, 125 (2005).
- [51] A. V. Hill, *J. Physiol.* **40**, iv (1910).
- [52] The assumption that such nonfundamental noise sources are negligible is supported by a detailed quantitative analysis of the transformation between Bicoid and Hunchback in the early *Drosophila* embryo [29], although it is not known whether this is true more generally.
- [53] M. B. Elowitz, M. G. Surette, P.-E. Wolf, J. B. Stock, and S. Leibler, *J. Bacteriol.* **181**, 197 (1999).
- [54] I. Golding and E. C. Cox, *Phys. Rev. Lett.* **96**, 098102 (2006).
- [55] J. Elf, G.-W. Li, and X. S. Xie, *Science* **316**, 1191 (2007).
- [56] This normalization of the mean expression levels is not exactly the same as fixing an absolute maximum, but in the small noise limit the difference is not significant.
- [57] P. V. Pedone, R. Ghirlando, G. M. Clore, A. M. Gronenborn, G. Felsenfeld, and J. G. Omichinski, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2822 (1996).
- [58] S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, and B. Müller-Hill, *EMBO J.* **13**, 3348 (1994).
- [59] Y. Wang, L. Guo, I. Golding, E. C. Cox, and N. P. Ong, *Biophys. J.* **96**, 609 (2009).
- [60]  $N_{\max}$  is the maximum number of independent molecular events, which is not the same as the maximum number of protein molecules expressed, but they are proportional.
- [61] Most transcription factors act either as activators or repressors, but not both. Thus, we should not take the existence of locally optimal solutions with “mixed” functions for the input molecule too literally. We think the important lesson here is that even simple formulations of our optimization problem have many solutions, so we expect that a realistic formulation, with multiple interacting targets, will also have multiple solutions. This is crucial if we want a theory based on optimization to predict the diversity of biological mechanisms.
- [62] Our discussion here has been in the limit where expression levels are in steady state. A complementary view of why tiling of the concentration axis is important is that, as the transcription factor concentration changes over time, the cell would cross the thresholds for activation or repression of different genes in sequence generating a temporal program of expression in which different proteins are synthesized at different times. See, for example, Refs. [73,74].
- [63] C. Waters and B. L. Bassler, *Genes Dev.* **20**, 2754 (2006).
- [64] V. Molle, M. Fujita, S. T. Jensen, P. Eichenberger, J. E. González-Pastor, J. S. Liu, and R. Losick, *Mol. Microbiol.* **50**, 1683 (2003).
- [65] M. Fujita, J. E. González-Pastor, and R. Losick, *J. Bacteriol.* **187**, 1357 (2005).
- [66] F. H. Lam, D. J. Steger, and E. K. O’Shea, *Science* **453**, 246 (2008).
- [67] W. Driever and C. Nüsslein-Vollhard, *Cell* **54**, 83 (1988).
- [68] W. Driever and C. Nüsslein-Vollhard, *Cell* **54**, 95 (1988).
- [69] W. Driever and C. Nüsslein-Vollhard, *Nature (London)* **337**, 138 (1989).
- [70] G. Struhl, K. Struhl, and P. M. Macdonald, *Cell* **57**, 1259 (1989).
- [71] D. S. Burz, R. Rivera-Pomar, H. Jäckle, and S. D. Hanes, *EMBO J.* **17**, 5998 (1998).
- [72] For a sampling of the evidence and models that describe these interactions, see Refs. [4,75–78].
- [73] M. Ronen, R. Rosenberg, B. Shraiman, and U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10555 (2002).
- [74] A. Zaslaver, A. E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M. G. Surette, and U. Alon, *Nat. Genet.* **36**, 486 (2004).
- [75] H. Jäckle, D. Tautz, R. Schuh, E. Seifert, and R. Lehmann, *Nature (London)* **324**, 668 (1986).
- [76] R. Rivera-Pomar and H. Jäckle, *Trends Genet.* **12**, 478 (1996).
- [77] L. Sanchez and D. Thieffry, *J. Theor. Biol.* **211**, 115 (2001).
- [78] D. Yu and S. Small, *Curr. Biol.* **18**, 868 (2008).