Optimizing information flow in small genetic networks. III. A self-interacting gene

Gašper Tkačik,[1,3,*] Aleksandra M. Walczak,[2,3,†] and William Bialek[3,‡]

[1]*Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria*
[2]*CNRS-Laboratoire de Physique Théorique de l'École Normale Supérieure, 24 rue Lhomond, 75005 Paris, France*
[3]*Joseph Henry Laboratories of Physics, Lewis–Sigler Institute for Integrative Genomics, Princeton University,
Princeton, New Jersey 08544, USA*

Living cells must control the reading out or "expression" of information encoded in their genomes, and this regulation often is mediated by transcription factors—proteins that bind to DNA and either enhance or repress the expression of nearby genes. But the expression of transcription factor proteins is itself regulated, and many transcription factors regulate their own expression in addition to responding to other input signals. Here we analyze the simplest of such self-regulatory circuits, asking how parameters can be chosen to optimize information transmission from inputs to outputs in the steady state. Some nonzero level of self-regulation is almost always optimal, with self-activation dominant when transcription factor concentrations are low and self-repression dominant when concentrations are high. In steady state the optimal self-activation is never strong enough to induce bistability, although there is a limit in which the optimal parameters are very close to the critical point.

I. INTRODUCTION

In order to function and survive in the world, cells must make decisions about the reading out or "expression" of genetic information. This happens when a bacterium makes more or less of an enzyme to exploit the variations in the availability of a particular type of sugar and when individual cells in a multicellular organism commit to particular fates during the course of embryonic development. In all such cases, the control of gene expression involves the transmission of information from some input signal to the output levels of the proteins encoded by the regulated genes. Although the notion of information transmission in these systems usually is left informal, the regulatory power that the system can achieve—the number of reliably distinguishable output states that can be accessed by varying the inputs—is measured, logarithmically, by the actual information transmitted, in bits [1,2]. Since relevant molecules often are present at relatively low concentrations, or even small absolute numbers, there are irreducible physical sources of noise that will limit the capacity for information transmission. Cells, thus, face a trade-off between regulatory power (in bits) and resources (in molecule numbers). What can cells do to maximize their regulatory power at fixed expenditure of resources? More precisely, what can they do to maximize information transmission with bounded concentrations of the relevant molecules?

We focus on the case of transcriptional regulation, where proteins—called transcription factors (TFs)—bind to sites along the DNA and modulate the rate at which nearby genes are transcribed into messenger RNA. Because many of the regulated genes themselves code for TF proteins, regulatory interactions form a network. The general problem of optimizing information flow through such regulatory networks is quite hard, and we have tried to break this problem into manageable pieces. Given the signal and noise characteristics of the regulatory interactions, cells can try to match the distribution of input transcription factor concentrations to these features of the regulatory network; even simple versions of this matching problem make experimentally testable predictions [3,4]. Assuming that this matching occurs, some regulatory networks still have more capacity to transmit information, and we can search for these optimal networks by varying both the topology of the network connections and the strengths of the interactions along each link in the network (the "numbers on the arrows" [5]). We have addressed this problem, first, in simple networks where a single input transcription factor regulates multiple noninteracting genes [6] and then in interacting networks where the interactions have a feedforward structure [7]. But real genetic regulatory networks have loops, and our goal here is to study the simplest such case, where a single input transcription factor controls a single self-interacting gene. Does feedback increase the capacity of this system to transmit information? Are self-activating or self-repressing genes more informative? Since networks with feedback can exhibit multistability or oscillation, and, hence, a nontrivial phase diagram as a function of the underlying parameters, where in this phase diagram do we find the optimal networks?

Autoregulation, both positive and negative, is one of the simplest and most commonly observed motifs in genetic regulatory networks [8–10] and has been the focus of a number of experiments and modeling studies (see, for example, Refs. [11,12]). A number of proposals have been advanced to explain its ubiquitous presence. Negative feedback (self-repression) can speed up the response of the genetic regulatory element [13] and can reduce the steady-state fluctuations in the output gene expression levels [14]. Positive feedback (self-activation), on the other hand, slows down the dynamics of gene expression and sharpens the response of a regulated

—————
[*]gtkacik@ist.ac.at
[†]awalczak@lpt.ens.fr
[‡]wbialek@princeton.edu

gene to its external input. Self-activating genes, thus, could threshold graded inputs, transforming them into discrete, almost "digital," outputs [15], allowing the cell to implement binary logical functions [16]. If self-activation is very strong, it can lead to multistability, or switchlike behavior of the response, so the genetic regulatory element can store the information for long periods of time [17–19]; such elements will also exhibit hysteretic effects. Weak self-activation, which does not cause multistability, has been studied less extensively but could play a role in allowing the cell to implement a richer set of input-output relations [10]. Alternatively, if the self-activating gene product can diffuse into neighboring nuclei of a multicellular organism, the sharpening effect of self-activation can compensate for the "blurring" of responses due to diffusion and, hence, open more possibilities for noise reduction through spatial averaging [20,21].

Many of the ideas about the functional role of autoregulation are driven by considerations of noise reduction. The physical processes by which the regulatory molecules find and bind to their regulatory sites on the DNA, the operation of the transcriptional machinery, itself subject to thermal fluctuations, and the unavoidable shot noise inherent in producing a small number of output proteins all contribute toward the stochastic nature of gene expression and, thus, place physical limits on the reliability of biological computation [22–25]. In the past decade the advance of experimental techniques has enabled us to measure the total noise in gene expression and sometimes parse apart the contributions of various molecular processes toward this "grand total" [26–34]. Progress has also been made in understanding that gene regulatory mechanisms beyond simple binding and unbinding of transcription factors, such as DNA looping, can control the noise at the molecular level [35,36]. With the detailed knowledge about noise in gene expression, we can revisit the original question and ask: Can both forms of autoregulation help mitigate the deleterious effects of noise on information flow through the regulatory networks and, if so, how?

All of our previous work in information transmission in transcriptional regulation has been in the steady-state limit. A similar approach was taken in Ref. [37], where the authors analyze information flow in elementary circuits, including feedback, but with different model assumptions about network topology and noise. More recently, de Ronde and colleagues have systematically reexamined the role of feedback regulation on the fidelity of signal transmission for time-varying Gaussian signals in cases where the (nonlinear) behavior of the genetic regulatory element can be linearized around some operating point [38]. They found that autoactivation increases gain-to-noise ratios for low-frequency signals, whereas autorepression yields an improvement for high-frequency signals. While many of the functions of feedback involve dynamics, as far as we know, all analyses of information transmission with dynamical signals resort to linear approximations. Here we return to the steady-state limit, where we can treat gene regulatory elements as fully nonlinear devices. While we hope that our analysis of the self-regulated gene is interesting in itself, we emphasize that our goal is to build intuition for the analysis of more general networks with feedback.
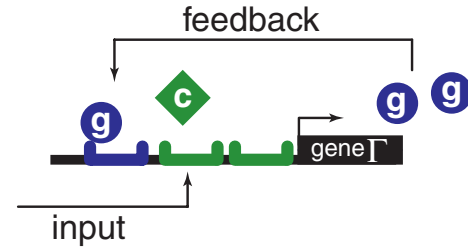

FIG. 1. (Color online) A schematic diagram of a self-regulating gene. The gene Γ is depicted by a thick black line and a promoter start signal. Gene products g denoted as blue circles can bind to the regulatory sites (one in this example) that control the expression of Γ. Direct control over the expression of Γ is exerted by molecules of the transcription factor c (green diamonds, two binding sites).

II. FORMULATING THE PROBLEM

Figure 1 shows a schematic of the system that we will analyze in this paper, a gene Γ that is controlled by two regulators: directly by an external transcription factor, as well as in a feedback fashion by its own gene products. We will refer to the transcription factor as the regulatory *input*; its concentration in the relevant (cellular or nuclear) volume Ω will be denoted by c. In addition, the gene products of Γ, whose number in the relevant volume Ω we denote by G and to which we refer to as the *output*, can also bind to the regulatory region of Γ, thereby activating or repressing the gene's expression. As we attempt to make our description of this system mathematically precise, the heart of our model will be the *regulatory function* that maps the concentrations of the two regulators at the promotor region of Γ to the rate at which output molecules are synthesized.

We can write the equation for the dynamics of gene expression from Γ by assuming that synthesis and degradation of the gene products are single kinetic steps, in which case we have

$$\frac{dG}{dt} = r_{\max} f(c,\gamma) - \frac{1}{\tau}G + \xi, \tag{1}$$

where $r_{\max}$ is the maximum rate for production of G, τ is the protein degradation time, and γ is the concentration of the output molecules in the relevant volume Ω. To include the noise effects inherent in creating and degrading single molecules of G we introduce the Langevin force ξ, and we will discuss the nature of this and other noise sources in detail later. Importantly, departures from our simplifying assumptions about the kinetics can, in part, be captured by proper treatment of the noise terms, as discussed below.

We are interested in the information that the steady-state output of Γ provides about the input concentration c. Following our previous work [6,7], we address this problem in stages. First we relate information transmission to the response properties and noise in the regulatory element, using a small-noise approximation to allow analytic progress (Sec. II A). We then show how the relevant noise variances can be computed from the model in Eq. (1), taking advantage of our understanding of the physics underlying the essential sources of noise (Sec. II B); this discussion is still quite general, independent of the details of the regulation function. We then explain our choice of the regulation function, adapted from

the Monod-Wyman-Changeux description of allosteric interactions (Sec. II C). Because feedback allows for bifurcations, we have to map the phase diagram of our model (Sec. II D) and develop approximations for the information transmission near the critical point (Sec. II E) and in the bistable regime (Sec. II F). Our discussion reviews some earlier results, in the interest of being self-contained, but the issues in Secs. II D–II F are all new to the case of networks with feedback.

A. Noise and information transmission

We are interested in computing the mutual information between the input and the output of a regulatory element in steady state. We have agreed that the input signal is the concentration c of the transcription factor, and we will take the output to be the concentration g of the gene products, which we colloquially call the expression level of the gene Γ. An important feature of the information transmission is that its mathematical definition is independent of the units that we use in measuring these concentrations, so when we later choose some natural set of units we will not have to worry about substituting into the formulas we derive here.

Following Shannon [1], the mutual information between c and g is defined by

$$I(c;g) = \int dc \int dg \, P(c,g) \log_2 \left[\frac{P(c,g)}{P_{\text{in}}(c) P_{\text{out}}(g)} \right] \text{ bits}, \quad (2)$$

where input concentrations c are drawn from the distribution $P_{\text{in}}(c)$, the output expression levels that we can observe are drawn from the distribution $P_{\text{out}}(g)$, and the joint distribution of these two quantities is $P(c,g)$. We think of the expression level as responding to the inputs, but this response will be noisy, so, given the input c, there is a conditional distribution $P(g|c)$. The symmetric expression for the mutual information in Eq. (2) then can be rewritten as a difference of entropies,

$$I(c;g) = S[P_{\text{out}}(g)] - \int dc \, P_{\text{in}}(c) S[P(g|c)], \quad (3)$$

where the entropy of a distribution is defined, as usual, by

$$S[P(x)] = -\int dx \, P(x) \log_2 P(x). \quad (4)$$

Finally, we recall that

$$P_{\text{out}}(g) = \int dc \, P_{\text{in}}(c) P(g|c). \quad (5)$$

Notice that the mutual information is a functional of two probability distributions, $P_{\text{in}}(c)$ and $P(g|c)$. The latter distribution describes the response and noise characteristics of the regulatory element and is something we will be able to calculate from Eq. (1). Following Refs. [3,4,6,7], we may then ask: Given that $P(g|c)$ is determined by the biophysical properties of the genetic regulatory element, what is the optimal choice of $P_{\text{in}}(c)$ that will maximize the mutual information $I(c;g)$? To this end we have to solve the problem of extremizing

$$\mathcal{L}[P_{\text{in}}(c)] = I(c;g) - \Lambda \int dc \, P_{\text{in}}(c), \quad (6)$$

where the Lagrange multiplier Λ enforces the normalization of $P_{\text{in}}(c)$. Other "cost" terms are possible, such as adding a term proportional to $\int dc \, c P_{\text{in}}(c)$, which would penalize the average cost of input molecules c, although here we take the simpler approach of fixing the maximum possible value of c, which is almost equivalent [6]. If the noise were truly zero, we could write the distribution of outputs as

$$P_{\text{out}}(g) = \int dc \, P_{\text{in}}(c) \delta[g - \bar{g}(c)], \quad (7)$$

where $\bar{g}(c)$ is the average output as a function of the input, i.e., the mean of the distribution $P(g|c)$. Then if the function $\bar{g}(c)$ is invertible, we can write the entropy of the output distribution as

$$S[P_{\text{out}}(g)] \equiv -\int dg \, P_{\text{out}}(g) \log_2 P_{\text{out}}(g)$$

$$\rightarrow -\int dc \, P_{\text{in}}(c) \log_2 \left[P_{\text{in}}(c) \left| \frac{d\bar{g}(c)}{dc} \right|^{-1} \right], \quad (8)$$

and we can think of this as the first term in an expansion in powers of the noise level [3]. Keeping only this leading term, we have

$$\mathcal{L}[P_{\text{in}}(c)] = -\int dc \, P_{\text{in}}(c) \log_2 \left[P_{\text{in}}(c) \left| \frac{d\bar{g}(c)}{dc} \right|^{-1} \right]$$

$$- \int dc \, P_{\text{in}}(c) S[P(g|c)] - \Lambda \int dc \, P_{\text{in}}(c), \quad (9)$$

and one can then show that the extremum of $\mathcal{L}$ occurs at

$$P_{\text{in}}^*(c) = \frac{1}{Z} 2^{-S[P(g|c)]} \left| \frac{d\bar{g}(c)}{dc} \right|, \quad (10)$$

where the entropy is measured in bits, as above, and the normalization constant,

$$Z = \int dc \left| \frac{d\bar{g}(c)}{dc} \right| 2^{-S[P(g|c)]}. \quad (11)$$

The maximal value of the mutual information is then simply $I^* = \log_2 Z$.

In the case where $P(g|c)$ is Gaussian,

$$P(g|c) = \frac{1}{\sqrt{2\pi \sigma_g^2(c)}} \exp \left\{ -\frac{[g - \bar{g}(c)]^2}{2\sigma_g^2(c)} \right\}, \quad (12)$$

the entropy is determined only by the variance $\sigma_g^2(c)$,

$$S[P(g|c)] = \frac{1}{2} \log_2 \left[2\pi e \sigma_g^2(c) \right]. \quad (13)$$

It is useful to think about propagating this (output) noise variance back through the input-output relation $\bar{g}(c)$, to define the effective noise at the input,

$$\sigma_c^2(c) = \left| \frac{d\bar{g}(c)}{dc} \right|^{-2} \sigma_g^2. \quad (14)$$

We then can write

$$P_{\text{in}}^*(c) = \frac{1}{Z} \frac{1}{\sqrt{2\pi e} \sigma_c(c)}. \quad (15)$$

As before, Z is the normalization constant,

$$Z = \int_0^C dc \left[\frac{1}{2\pi e} \left(\frac{d\bar{g}}{dc} \right)^2 \frac{1}{\sigma_g^2(c)} \right]^{1/2}, \quad (16)$$

where C is the maximal value of the input concentration, and, again, we have the information $I^*(c;g) = \log_2 Z$ bits.

B. Noise variances

Equation (16) relates Z, and, hence, the information transmission $I^* = \log_2 Z$, to the steady-state response and noise in our simple regulatory element. These quantities are calculable from the dynamical model in Eq. (1) using the Lagevin framework [39] and the Itō formulation of stochastic calculus that we follow in this manuscript, if we understand the sources of noise. There are two very different kinds of noise that we need to include in our analysis.

First, we are describing molecular events that synthesize and degrade individual molecules, and individual molecules behave randomly. If we say that there is synthesis of $\bar{r}$ molecules per second on average, then, if the synthesis is limited by a single kinetic step, and if all molecules behave independently, then the actual rate $r(t)$ will fluctuate with a correlation function $\langle \delta r(t) \delta r(t') \rangle = \bar{r} \delta(t - t')$. Similarly, if, on average, there is degradation of $\bar{d}$ molecules per second, then the actual degradation rate $d(t)$ will fluctuate with $\langle \delta d(t) \delta d(t') \rangle = \bar{d} \delta(t - t')$. Thus, if we want to describe the time dependence of the number of molecules $N(t)$, we can write,

$$\frac{dN}{dt} = r(t) - d(t) = \bar{r} - \bar{d} + \xi(t), \qquad (17)$$

where

$$\xi(t) = \delta r(t) - \delta d(t). \qquad (18)$$

If we are close to the steady state, $\bar{r} = \bar{d}$, and if synthesis and degradation reactions are independent, we have

$$\langle \xi(t) \xi(t') \rangle = 2\bar{d} \delta(t - t'). \qquad (19)$$

If some of the reactions involve multiple kinetic steps, or if the molecules we are counting are amplified copies of some other molecules, then the noise will be proportionally larger or smaller, and we can take account of this by introducing a "Fano factor" ν, so

$$\langle \xi(t) \xi(t') \rangle \rightarrow 2\nu \bar{d} \delta(t - t'). \qquad (20)$$

The second irreducible source of noise is that the synthesis reactions are regulated by transcription factor binding to DNA, and these molecules arrive randomly at their targets. One way to think about this is that the concentrations of TFs which govern the synthesis rate are not the bulk average concentrations over the whole cell or nucleus but rather concentrations in some small "sensitive volume" determined by the linear size ℓ of the targets themselves [22,40,41]. Concretely, if we write the synthesis rate as

$$r = r_{\max} f(\hat{c}, \gamma), \qquad (21)$$

where $\hat{c}$ is the local concentration of the input transcription factor and γ is the concentration of the gene product that feeds back to regulate itself, we should really think of these concentrations as $\hat{c} = c + \xi_c$ and $\gamma = G/\Omega + \xi_g$, where we separate the mean values and the local fluctuations; note

that the mean gene product concentration is the ratio of the molecule number G to the relevant volume Ω. The local concentration fluctuations are also white, and the spectral densities are given accurately by dimensional analysis [22,40], so

$$\langle \xi_c(t) \xi_c(t') \rangle = (2c/D\ell)\delta(t - t'), \qquad (22)$$

$$\langle \xi_\gamma(t) \xi_\gamma(t') \rangle = (2G/\Omega D\ell)\delta(t - t'), \qquad (23)$$

where D is the diffusion constant of the transcription factor molecules, which we assume is the same for the input and output proteins.

We can put all of these factors together if the noise is small, so that it drives fluctuations which stay in the linear regime of the dynamics. Then, if the steady-state solution to Eq. (1) in the absence of noise is denoted by $G = \bar{G}(c)$, we can linearize in the fluctuations $\delta G = G - \bar{G}$ as follows:

$$\frac{dG}{dt} = r_{\max} f(\hat{c}, \gamma) - \frac{1}{\tau} G + \xi$$

$$= r_{\max} f(c + \xi_c, G/\Omega + \xi_\gamma) - \frac{1}{\tau} G + \xi, \qquad (24)$$

$$\Rightarrow \frac{d(\delta G)}{dt} = \left[\frac{r_{\max}}{\Omega} \left. \frac{\partial f(c, \gamma)}{\partial \gamma} \right|_{\gamma = \bar{G}/\Omega} - \frac{1}{\tau} \right] \delta G + \xi_{\text{eff}}(t), \qquad (25)$$

where

$$\langle \xi_{\text{eff}}(t) \xi_{\text{eff}}(t') \rangle = 2 \left[\nu \frac{\bar{G}}{\tau} + \left(r_{\max} \left. \frac{\partial f(c, \gamma)}{\partial \gamma} \right|_{\gamma = \bar{G}/\Omega} \right)^2 \frac{2\bar{G}}{\Omega D\ell} \right.$$
$$\left. + \left(r_{\max} \frac{\partial f(c, \bar{G}/\Omega)}{\partial c} \right)^2 \frac{2c}{D\ell} \right] \delta(t - t'). \qquad (26)$$

To solve this problem and compute the variance in the output number of molecules $\langle (\delta G)^2 \rangle$, it is useful to recall the Langevin equation for the position $x(t)$ of an overdamped mass tethered by a spring of stiffness κ, subject to a drag force proportional to the velocity, $F_{\text{drag}} = -\alpha v$:

$$\alpha \frac{dx}{dt} = -\kappa x + \zeta(t), \qquad (27)$$

$$\langle \zeta(t) \zeta(t') \rangle = 2k_B T \alpha \delta(t - t'). \qquad (28)$$

From equipartition we know that these dynamics predict the variance $\langle x^2 \rangle = k_B T/\kappa$. Identifying terms with our Langevin description of the synthesis and degradation reactions, we find

$$\langle (\delta G)^2 \rangle = \frac{1}{\frac{1}{\tau} - \frac{r_{\max}}{\Omega} \frac{\partial f}{\partial \gamma}} \left[\nu \frac{\bar{G}}{\tau} + \left(r_{\max} \frac{\partial f}{\partial \gamma} \right)^2 \frac{\bar{G}}{\Omega D\ell} \right.$$
$$\left. + \left(r_{\max} \frac{\partial f}{\partial c} \right)^2 \frac{c}{D\ell} \right], \qquad (29)$$

where we understand that the partial derivatives of f are to be evaluated at the steady state $\gamma = \bar{G}/\Omega$.

We have defined $r_{\max}$ as the maximum synthesis rate, so the regulation function f is in the range $0 \leqslant f \leqslant 1$ and, hence, the maximum mean expression level is $\bar{G}_{\max} = r_{\max}\tau$. Thus, it makes sense to work with a normalized expression level $g \equiv G/(r_{\max}\tau)$, and to think of the regulation function f as depending on g rather than on the absolute concentration γ. We then have

$$\langle (\delta g)^2 \rangle \equiv \sigma_g^2(c) = \frac{1}{1 - (\partial f/\partial g)} \left[\left(\frac{\nu}{r_{\max}\tau} \right) \bar{g} + \left(\frac{\partial f}{\partial g} \right)^2 \bar{g} \frac{1}{D\ell\gamma_{\max}} + \left(\frac{\partial f}{\partial c} \right)^2 \frac{c}{D\ell} \right], \quad (30)$$

where $\gamma_{\max}$ is the maximal mean concentration of output molecules. As discussed previously [6], we can think of $N_g = r_{\max}\tau/\nu$ as the maximum number of independent output molecules, and this combines with the other parameters in the problem to define a natural concentration scale, $c_0 = N_g/(D\ell\tau)$. Once we choose units where $c_0 = 1$, we have a simpler expression,

$$\sigma_g^2(c) = \frac{1}{N_g} \frac{1}{1 - (\partial f/\partial g)} \left[\bar{g} + \left(\frac{\partial f}{\partial g} \right)^2 \frac{\bar{g}}{\gamma_{\max}} + \left(\frac{\partial f}{\partial c} \right)^2 c \right], \quad (31)$$

where we notice that almost all the parameters have been eliminated by our choice of units.

Finally, we need to use the variance to compute the information capacity of the system, $I^* = \log_2 Z$, where, from Eq. (16), we have

$$Z = \int_0^C dc \left[\frac{1}{2\pi e} \left(\frac{d\bar{g}}{dc} \right)^2 \frac{1}{\sigma_g^2(c)} \right]^{1/2} = \left[\frac{N_g}{2\pi e} \right]^{1/2} \tilde{Z}, \quad (32)$$

$$\tilde{Z} = \int_0^C dc \left[\left(\frac{d\bar{g}}{dc} \right)^2 \frac{1 - (\partial f/\partial g)}{\bar{g} + (\partial f/\partial g)^2 (\bar{g}/\gamma_{\max}) + (\partial f/\partial c)^2 c} \right]^{1/2}, \quad (33)$$

where C is the maximum concentration of input transcription factor molecules, in units of c_0. Notice that the parameter N_g just scales the noise and (in the small-noise approximation) thus adds to the information, $I^* = \log_2 \tilde{Z} + (1/2) \log_2(N_g/2\pi e)$; the problem of optimizing information transmission, thus, is the problem of optimizing $\tilde{Z}$. Further, because

$$\bar{g} = f(c, \bar{g}), \quad (34)$$

the total derivative $d\bar{g}/dc$ can be expressed though

$$\frac{d\bar{g}}{dc} = \frac{\partial f}{\partial c} + \frac{\partial f}{\partial g} \frac{d\bar{g}}{dc}. \quad (35)$$

Putting all of these pieces together, we find

$$\tilde{Z} = \int_0^C dc \frac{\frac{\partial f}{\partial c} \left(1 - \frac{\partial f}{\partial g} \right)^{-\frac{1}{2}}}{\sqrt{\bar{g} + \left(\frac{\partial f}{\partial c} \right)^2 c + \left(\frac{\partial f}{\partial g} \right)^2 \bar{g}/\gamma_{\max}}}. \quad (36)$$

In what follows, we will start with the assumption that, since both input and output molecules are transcription factor proteins, their maximal concentrations are the same, and, hence, $\gamma_{\max} = C$; we will return to this assumption at the end of our discussion.

If a regulatory function $f(c,g)$ is chosen from some parametric family, Eq. (36) allows us to compute the information transmission as a function of these parameters and search for an optimum. Before embarking on this path, however, we note that the integrand of $\tilde{Z}$ can have a divergence if $\partial f/\partial g = 1$. This is a condition for the existence of a *critical point*, and in this simple system the critical point or bifurcation separates the regime of monostability from the regime of bistability. We expect that, at this point, the fluctuations around $\bar{g}$ are no longer Gaussian, and we need to compute higher-order moments. Thus, Eq. (36), as is, can safely be used only in the monostable regime away from the critical point; in Sec. II E we compute the expression for the mutual information near to and at the critical point for a particular choice of f. There are even more problems in the bistable regime, since there are multiple solutions to Eq. (34), and in Sec. II F we discuss information in the bistable regime.

C. MWC regulatory function

To continue, we must choose a regulatory function. In Ref. [7], where we analyzed genetic networks with feed-forward interactions, we studied Hill-type regulation [42] and Monod-Wyman-Changeux–like (MWC) regulation [43] and found that the MWC family encompasses a broader set of functions than the Hill family; for a related discussion see Ref. [44]. MWC functions also allow for a natural introduction of convergent control, where a node in a network is simultaneously regulated by several types of regulatory molecules. Briefly, in the MWC model, one assumes that the molecule or supermolecular complex being considered has two states, which we identify here with ON and OFF states of the promoter. The binding of each different regulatory factor is always independent, but the binding energies depend on whether the complex is in an OFF or ON state, so (by detailed balance) binding shifts the equilibrium between these two states.

In our case, we have two regulatory molecules, the input transcription factor with concentration c and the gene product with concentration g. If there are, respectively, n_c and n_g identical binding sites for these molecules, then the probability of being in the ON state is

$$f(c,g) = \frac{\left(1 + c/Q_c^{\mathrm{on}} \right)^{n_c} \left(1 + g/Q_g^{\mathrm{on}} \right)^{n_g}}{L \left(1 + c/Q_c^{\mathrm{off}} \right)^{n_c} \left(1 + g/Q_g^{\mathrm{off}} \right)^{n_g} + \left(1 + c/Q_c^{\mathrm{on}} \right)^{n_c} \left(1 + g/Q_g^{\mathrm{on}} \right)^{n_g}}, \quad (37)$$

where $Q_c^{\mathrm{on}}, Q_g^{\mathrm{on}}$ are the binding constants in the ON state and, similarly, $Q_c^{\mathrm{off}}, Q_g^{\mathrm{off}}$ are the binding constants in the OFF state; L reflects the "bare" free energy difference between the two states. If the binding of the regulatory molecules has a strong activating effect, then we expect $Q_c^{\mathrm{on}} \ll Q_c^{\mathrm{off}}$, and similarly for Q_g, which means that only one binding constant is relevant for each molecule, and we will refer to these as K_c and K_g. Then we can write

$$f(c,g) = \frac{1}{1 + e^{-F(c,g)}}, \tag{38}$$

$$F(c,g) = n_c \ln\left(\frac{1 + c/K_c}{1 + c_{1/2}/K_c}\right) + n_g \ln\left[\frac{1 + g/K_g}{1 + 1/(2K_g)}\right], \tag{39}$$

where $c_{1/2}$ is the input concentration at which $\bar{g}(c_{1/2}) = 1/2$. Notice that if binding of g strongly represses the gene, then we have $Q_g^{\mathrm{off}} \ll Q_g^{\mathrm{on}}$, but this can be simulated by changing the sign of n_g. Thus, we should think of the parameters n_c and n_g as being not just the number of binding sites but also an index of activation versus repression. We will also treat these parameters as continuous, which is a bit counterintuitive but allows us to describe, approximately, situations in which the multiple binding sites are inequivalent or in which Q^{on} and Q^{off} do not infinitely differ.

From the discussion in the previous section, we will need to evaluate the partial derivatives of $f(c,g)$ with respect to its arguments. For the MWC model, these derivatives take simple forms:

$$\frac{\partial f}{\partial c} = f(1 - f)\frac{n_c}{K_c + c}, \tag{40}$$

$$\frac{\partial f}{\partial g} = f(1 - f)\frac{n_g}{K_g + g}. \tag{41}$$

D. Phase diagram

Let us start by examining the stability properties of Eq. (34), which determines the steady state $\bar{g}(c)$. Viewed as a function of g, $f(c,g)$ is sigmoidal, and so, if we try to solve $g = f(c,g)$ graphically, we are looking for the intersection of a sigmoid with the diagonal as a function of g. In doing this, we expect that, for some values of the parameters, there will be exactly one solution but that, as we change parameters (or the input c), there will be a transition to multiple solutions. This transition happens when f just touches the diagonal, that is, when for some $\bar{g}^*$ it holds true that $f(\bar{g}^*, c) = \bar{g}^*$ and $\partial f(\bar{g}, c)|_{\bar{g}^*}/\partial \bar{g} = 1$. Using Eq. (41), these two conditions can be combined to yield an equation for $\bar{g}^*$:

$$\bar{g}^*(1 - \bar{g}^*)\frac{n_g}{K_g + \bar{g}^*} = 1. \tag{42}$$

This is a quadratic in $\bar{g}^*$ for which no real solution on $\bar{g}^* \in [0,1]$ exists if either

$$K_g > \frac{(n_g - 1)^2}{4 n_g} \tag{43}$$

or $n_g < 1$. When either of these conditions are fulfilled, the gene is in the monostable regime. At the critical point,
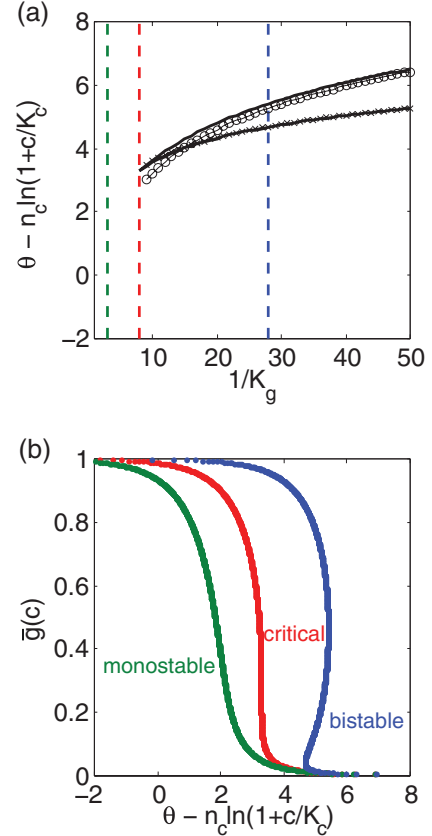

FIG. 2. (Color online) Monostable, critical, and bistable (green, red, and blue lines, left to right, respectively) behavior of the self-activating gene for $n_g = 2$. (a) The phase diagram as a function of K_g and the input-dependent term. In the region between the black solid lines two solutions $\bar{g}_{1,2}$ exist for every value of the input c (y axis). The corresponding critical value is at $K_g^* = 1/8$ (cusp of the black solid lines). Circles and crosses represent analytical approximations to the exact boundary of the bistable region for $n_g = 2$ and large $1/K_g$ (see text). For three choices of K_g denoted by vertical dashed lines, the input-output relations $\bar{g}(c)$ are plotted in (b). (b) The critical solution (red) has an infinite total derivative $\bar{g}'(c)$ at $\theta - n_c \log(1 + c/K_c) = 3.29$, $\bar{g} = 0.25$. The bistable system (blue) has three solutions, two stable and one unstable, for a range of inputs that can be read out from the plot in (a).

$K_g^* = (n_g - 1)^2/(4 n_g)$ and $\bar{g}^* = (n_g - 1)/(2 n_g)$. This is illustrated in Fig. 2 as a function of the effective input $\chi(c) = \theta - n_c \ln(1 + c/K_c)$, where, from Eq. (39), $\theta = n_c \ln(1 + c_{1/2}/K_c) + n_g \ln[1 + 1/(2K_g)]$.

For the special case of $n_g = 2$ it is not hard to compute the analytical approximations for the boundary of the bistable domain. First, Eq. (38) can be expanded for large K_g, yielding a quadratic equation for $\bar{g}$ that has two solutions only when

$$\chi < -\ln 4 - 2 \ln K_g. \tag{44}$$

To get the lower bound, we expand Eq. (38) for small $\bar{g}$ and retain terms up to the quadratic order in $\bar{g}$; the resulting quadratic equation yields two solutions only if

$$\chi > \ln(4/K_g - 3). \tag{45}$$

Both approximations are plotted as circles and crosses, respectively, in Fig. 2, and match the exact curves well. For other

values of n_g we solve Eq. (34) exactly, using a bisection method to get all solutions for a given c and we partition the range of c adaptively into a denser grid where the derivative $\bar{g}'(c)$ is large. For integer values of n_g when the equation can be rewritten as a polynomial in $\bar{g}$, it is technically easier to find the roots of the polynomial; alternatively, one can solve for c given $\bar{g}$ using a simple bisection, because $c(\bar{g})$ is an injective function.

E. Information transmission near the critical point

In this section we will generalize the computation of noise and information in the region close to the critical point, where the Gaussian noise approximation breaks down. We start by rewriting Eqs. (24) and (25) in our normalized units,

$$\tau \frac{dg}{dt} = f(c,g) - g + \zeta, \tag{46}$$

$$\langle \zeta(t)\zeta(t') \rangle = 2\tau T(g)\delta(t - t'), \tag{47}$$

$$T(g) = \frac{1}{N_g}\left[g + \left(\frac{\partial f}{\partial g} \right)^2 \frac{g}{\gamma_{\max}} + \left(\frac{\partial f}{\partial c} \right)^2 c \right]. \tag{48}$$

This is equivalent to Brownian motion of the coordinate g in a potential $V(g,c)$ defined by

$$-\frac{dV(g,c)}{dg} = f(c,g) - g, \tag{49}$$

with an effective temperature $T(g)$ that varies with position. If we simulate this Langevin equation, we will draw samples out of the distribution $P(g|c)$, but we can construct this distribution directly by solving the equivalent diffusion or Fokker-Planck equation,

$$\frac{\partial P(g,t)}{\partial t} = \frac{\partial}{\partial g}[V'(g,c)P(g,t)] + \frac{\partial^2}{\partial g^2}[T(g)P(g,t)]. \tag{50}$$

The steady-state solution is then $P(g|c)$, and this is

$$P(g|c) = P(0)\frac{T(0)}{T(g)}\exp\left[-\int_0^g \frac{V'(y,c)}{T(y)}dy \right]. \tag{51}$$

The "small-noise approximation" in this extended framework corresponds to expanding the integrand in Eq. (51) around the mean, $g = \bar{g}(c) + \delta g$. If we write $x = g - \bar{g}(c)$, we will find

$$P(g|c) \propto \exp(-a_2 x^2 - a_3 x^3 - a_4 x^4 - \cdots), \tag{52}$$

where our previous approximations correspond to keeping only a_2. The critical point is where $a_2 = 0$, and we have to keep higher-order terms. In principle, the expansion coefficients have contributions from the g dependence of the effective temperature, but we have checked that these contributions are negligible near criticality. We then have

$$a_2 = \frac{1}{2!}\left[\frac{1 - f'}{T} \right], \quad a_3 = -\frac{1}{3!}\frac{f''}{T}, \quad a_4 = -\frac{1}{4!}\frac{f'''}{T}, \tag{53}$$

where primes denote derivatives with respect to g, and all terms should be evaluated at $g = \bar{g}(c)$.

For the Monod-Wyman-Changeux regulatory function in Eq. (38), all these derivatives can be evaluated explicitly,

$$f' = \frac{\partial f}{\partial g} = f(1 - f)\frac{n_g/K_g}{1 + g/K_g}, \tag{54}$$

$$f'' = \frac{\partial^2 f}{\partial g^2} = f(1 - f)\left(\frac{n_g/K_g}{1 + g/K_g} \right)^2\left(1 - 2f - \frac{1}{n_g} \right), \tag{55}$$

$$f''' = \frac{\partial^3 f}{\partial g^3} = f(1 - f)\left(\frac{n_g/K_g}{1 + g/K_g} \right)^3\left[\left(1 - 2f - \frac{1}{n_g} \right) \right.$$
$$\left. \times \left(1 - 2f - \frac{2}{n_g} \right) - 2f(1 - f) \right]. \tag{56}$$

From Eq. (42), the critical point occurs at $\bar{g}^* = (n_g - 1)/2n_g$ when $K_g^* = (n_g - 1)^2/4n_g$, and at this point the derivatives simplify as follows:

$$f' = 1, \tag{57}$$

$$f'' = 0, \tag{58}$$

$$f''' = -\frac{8n_g^2}{n_g^2 - 1}. \tag{59}$$

Now we want to explore behavior in the vicinity of the critical point; we will fix K_g to its critical value, $K_g^*(n_g)$, and compute the derivatives in Eqs. (54)–(56) as $\partial f/\partial g \to 1$. Consider, therefore, a small positive ϵ such that

$$\left. \frac{\partial f}{\partial g} \right|_{\bar{g}} = 1 - \epsilon. \tag{60}$$

In a system with chosen K_g and n_g that yield critical behavior, the deviation from criticality above will happen at $\bar{g} = \bar{g}^* + \Delta$. To find the relation between ϵ and Δ, we evaluate the derivative in Eq. (54) at $\bar{g}$ to form a function $\psi(\bar{g}) = \bar{g}(1 - \bar{g})n_g/(K_g + \bar{g})$, which evaluates to 1 at $\bar{g}^*$. This function can be expanded in Taylor series around $\bar{g}^*$; the first order in Δ vanishes and we find,

$$\frac{n_g(K_g + 1)}{(1 + \bar{g}^*/K_g)^3}\Delta^2 = \epsilon. \tag{61}$$

Therefore, the derivative deviates by ϵ from criticality at 1 when $\bar{g}$ deviates by $\pm\Delta$ from the $\bar{g}^*$. We now perform similar expansions on the second- and third-order derivatives in Eqs. (55) and (56) and evaluate the factors at the critical point,

$$\frac{\partial f}{\partial g} = 1 - \frac{4n_g^2}{n_g^2 - 1}\Delta^2, \tag{62}$$

$$\frac{\partial^2 f}{\partial g^2} = -\frac{8n_g^2}{n_g^2 - 1}\Delta + \frac{32n_g^3}{(n_g^2 - 1)^2}\Delta^2, \tag{63}$$

$$\frac{\partial^3 f}{\partial g^3} = -\frac{8n_g^2}{n_g^2 - 1} + \frac{64n_g^3}{(n_g^2 - 1)^2}\Delta + \frac{128n_g^4(n_g^2 - 4)}{(n_g^2 - 1)^3}\Delta^2. \tag{64}$$

These expressions have been evaluated for K_g^*, but we could have easily repeated the calculation by assuming that K_g itself can deviate a bit from the critical value, i.e., $K_g = K_g^* + \delta K_g$, which would yield somewhat more complicated results that we do not reproduce here.

Equations (62)–(64) can be used in Eq. (53) to write the probability distribution $P(g|c)$. Far from the critical point, the Gaussian approximation is assumed to hold, and a_3, a_4 can be set to 0. Close to the critical point, the higher-order terms a_3 and a_4 need to be included. To assess the range where this

switchover occurs, we compare in Eqs. (62)–(64) the leading to the subdominant correction: We insist that the quadratic correction in Eq. (63) is always smaller than linear and that the linear correction in Eq. (64) is always smaller than constant (we drop the quadratic correction there). We found empirically that including the higher-order corrections yields good results when the following conditions are simultaneously satisfied:

$$|\Delta| < \frac{n_g^2 - 1}{16 n_g} \quad |\Delta| < 0.25. \quad (65)$$

These conditions guarantee that the higher-order terms, which are evaluated around the critical point, are nowhere evaluated too far from the critical point such that the approximations would break down.

We can now put the pieces together by using the general form of the optimal solution for $P^*(c)$ in Eq. (10), together with the quartic ansatz for $P(g|c)$ in Eq. (52). For each c, we evaluate two entropies of the conditional distribution $P(g|c)$:

$$S_2[P(g|c)] = \log_2 \sqrt{2\pi e \sigma_g^2(c)}, \quad (66)$$

$$S_4[P(g|c)] = -\int dg \, P(g|c) \log_2 P(g|c). \quad (67)$$

S_4 is the noise entropy with higher-order terms included whenever conditions Eq. (65) are met, and S_2 is the noise entropy in the Gaussian approximation.

Equation (10) can be rewritten in a numerically stable fashion by realizing that $P^*(c)|d\bar{g}/dc|^{-1} = P^*(\bar{g})$, that is, that the optimal distribution of mean output levels is given by

$$P^*(\bar{g}) = 2^{-S[P(g|\bar{g})]}/Z. \quad (68)$$

To join the Gaussian and higher-order approximations consistently in the regimes away and near the critical point, the noise entropy in Eq. (69) is chosen to be the pointwise minimum of $S_4(\bar{g})$ and $S_2(\bar{g})$. Finally, the information is, again, $I = \log_2 Z$, with

$$Z = \int_0^{\bar{g}(C)} d\bar{g} \, 2^{-S[P(g|\bar{g})]}. \quad (69)$$

F. Information transmission in the bistable regime

We now discuss the information capacity in the bistsable regime, away from the critical line. In this regime, each value of the input c can give rise to multiple solutions of the steady-state equation (34). In the simplest case (which includes the MWC regulatory functions), there will be two stable solutions, $\bar{g}_1(c)$ and $\bar{g}_2(c)$, and a third solution, $\bar{g}_3(c)$, that is unstable. In equilibrium, the system will be on the first branch with weight $w_1(c)$ and on the second with weight $w_2(c)$. Here we place upper bound on the information $I(c; g)$, again, in the small-noise approximation. This will be useful since, as we will see, even this upper bound is always less than the information which can be transmitted in the monostable or critical regime, and so we will be able to conclude that the optimal parameters for which we are searching are never in the bistable regime.

In the bistable regime, the small-noise approximation (again, away from the critical line) means that the conditional distributions are well approximated by a mixture of Gaussians,

$$P(g|c) = w_1(c) \frac{1}{\sqrt{2\pi \sigma_1^2(c)}} e^{-\frac{(g-\bar{g}_1(c))^2}{2\sigma_1^2(c)}}$$

$$+ w_2(c) \frac{1}{\sqrt{2\pi \sigma_2^2(c)}} e^{-\frac{(g-\bar{g}_2(c))^2}{2\sigma_2^2(c)}}. \quad (70)$$

To compute the information we need two terms, the total entropy and the conditional entropy. The conditional entropy takes a simple form if we assume the noise is small enough that the Gaussians do not overlap. A direct calculation then shows that, as one might expect intuitively, the conditional entropy is just the weighted sum of the entropies of the Gaussian distributions plus a term that reflects the uncertainty about which branch the system is on,

$$S[P(g|c)] = \frac{1}{2} \sum_{i=1}^{2} w_i(c) \log_2 \left[2\pi e \sigma_i^2(c) \right]$$

$$- \sum_{i=1}^{2} w_i(c) \log_2 w_i(c). \quad (71)$$

Implementing the small-noise approximation for the total entropy is a bit more subtle. We have, as usual,

$$P(g) \equiv \int dc \, P(c) P(g|c)$$

$$= \int dc \, P(c) \frac{w_1(c)}{\sqrt{2\pi \sigma_1^2(c)}} \exp\left\{ -\frac{[g - \bar{g}_1(c)]^2}{2\sigma_1^2(c)} \right\}$$

$$+ \int dc \, P(c) \frac{w_2(c)}{\sqrt{2\pi \sigma_2^2(c)}} \exp\left\{ -\frac{[g - \bar{g}_2(c)]^2}{2\sigma_2^2(c)} \right\}. \quad (72)$$

If the noise is small, each of the two integrals is dominated by values of c near the solution of the equation $g = g_i(c)$; let us call these solutions $\hat{c}_i(g)$. Notice that these solutions might not exist over the full range of g, depending on the structure of the branches. Nonetheless, we can write,

$$P(g) \approx \left[w_1(c) P(c) \left| \frac{d\bar{g}_1(c)}{dc} \right|^{-1} \right]_{c=\hat{c}_1(g)}$$

$$+ \left[w_2(c) P(c) \left| \frac{d\bar{g}_2(c)}{dc} \right|^{-1} \right]_{c=\hat{c}_2(g)}, \quad (73)$$

with the convention that if we try to evaluate $w_i(c)$ at a nonexistent value of $\hat{c}_i$, we get zero. Thus, the full distribution $P(g)$ is also a mixture,

$$P(g) = f_1 P_1(g) + f_2 P_2(g). \quad (74)$$

The fractional contributions of the two distributions are

$$f_i = \int_i dc \, P(c) w_i(c), \quad (75)$$

where $\int_i dc \cdots$ denotes an integral over the regions along the c axis where the function $\hat{c}_i(g)$ exists, and the (normalized)

component distributions are

$$P_i(g) = \frac{1}{f_i} \left[w_i(c) P(c) \left| \frac{d\bar{g}_i(c)}{dc} \right|^{-1} \right]_{c = \hat{c}_i(g)}. \tag{76}$$

The entropy of a mixture is always less than the average entropy of the components, so we have an upper bound,

$$S[P(g)] \leqslant - \sum_{i=1}^{2} f_i \int dg \, P_i(g) \log_2 P_i(g) \tag{77}$$

$$= - \sum_{i=1}^{2} \int dg \left[w_i(c) P(c) \left| \frac{d\bar{g}_i(c)}{dc} \right|^{-1} \right]_{c = \hat{c}_i(g)} \log_2 \left[\frac{1}{f_i} w_i(c) P(c) \left| \frac{d\bar{g}_i(c)}{dc} \right|^{-1} \right]_{c = \hat{c}_i(g)} \tag{78}$$

$$= - \sum_{i=1}^{2} \int_i dc \, P(c) w_i(c) \log_2 \left[\frac{1}{f_i} w_i(c) P(c) \left| \frac{d\bar{g}_i(c)}{dc} \right|^{-1} \right]. \tag{79}$$

An upper bound on the total entropy is useful because it allows us to bound the mutual information:

$$I(c;g) \equiv S[P(g)] - \int dc \, P(c) S[P(g|c)] \tag{80}$$

$$\leqslant - \sum_{i=1}^{2} \int_i dc \, P(c) w_i(c) \log_2 \left[\frac{1}{f_i} w_i(c) P(c) \left| \frac{d\bar{g}_i(c)}{dc} \right|^{-1} \right]$$

$$- \frac{1}{2} \sum_{i=1}^{2} \int dc \, P(c) w_i(c) \log_2 \left[2\pi e \sigma_i^2(c) \right] + \sum_{i=1}^{2} \int dc \, P(c) w_i(c) \log_2 w_i(c) \tag{81}$$

$$= - \int dc \, P(c) \log_2 P(c) + \frac{1}{2} \int dc \, P(c) \sum_{i=1}^{2} w_i(c) \log_2 \left[\left| \frac{d\bar{g}_i(c)}{dc} \right|^2 \frac{1}{2\pi e \sigma_i^2(c)} \right] - \left(- \sum_{i=1}^{2} f_i \log_2 f_i \right) \tag{82}$$

$$\leqslant - \int dc \, P(c) \log_2 P(c) + \frac{1}{2} \int dc \, P(c) \sum_{i=1}^{2} w_i(c) \log_2 \left[\left| \frac{d\bar{g}_i(c)}{dc} \right|^2 \frac{1}{2\pi e \sigma_i^2(c)} \right], \tag{83}$$

where, in the last step, we use the positivity of the entropy associated with the mixture weights $\{f_i\}$.

We can now ask for the probability distribution $P(c)$ that maximizes the upper bound on $I(c;g)$, and, in this way, we can bound the capacity of the system. Happily, the way in which the bound depends on $P(c)$, in Eq. (83), does not differ so much from the dependencies that we have seen in the monostable case Eq. (9), so we can follow a parallel calculation to show that

$$I(g;c) \leqslant \log_2 \left[\int_0^C dc \, e^{-\phi(c)} \right], \tag{84}$$

$$\phi(c) = \sum_i w_i(c) \ln \left[\sqrt{2\pi e \sigma_{g_i}^2(c)} \left| \frac{d\bar{g}_i(c)}{dc} \right|^{-1} \right]. \tag{85}$$

Finally, to find the weights $w_i(c)$ we can numerically integrate the Fokker-Planck solution in Eq. (51) to find

$$w_1(c) = \int_0^{\bar{g}_3(c)} dg' \, P(g'|c), \tag{86}$$

$$w_2(c) = \int_{\bar{g}_3(c)}^{\infty} dg' \, P(g'|c). \tag{87}$$

To summarize, we have derived an upper bound on the information transmitted between the input and the output. The tightness of the bound is related to the applicability of the "no overlap" approximation, which for MWC-like regulatory functions should hold very well, as we have verified numerically. If only one of the weights $w_i \neq 0$, our results reduce to those in the monostable case, as they should.

III. RESULTS

We begin by showing that the analytical calculations presented in the previous section can be carried out numerically in a stable fashion, both away from and in the critical regime. We recall that the information transmission is determined by an integral $\tilde{Z}$ [Eq. (36)] and that, because we are working in the small-noise approximation, we have a choice of evaluating this as an integral over the input concentration c or an integral over the mean output concentrations $\bar{g}$. Figure 3 shows the behavior of the integrands in these two equivalent formulations when we have chosen parameters that are close to the critical point in a self-activating gene. The key result is that, once we include terms beyond the Gaussian approximation to $P(g|c)$ following the discussion in Sec. II E, we really do have control over the calculation and find smooth results as the parameter values approach criticality. Thus, we can compute confidently, and search for parameters of the regulatory function $\theta \equiv \{c_{1/2}, K_c, n_c, K_g, n_g\}$ that maximize information transmission.

We start the optimization by choosing the parameter values $\theta_g \equiv \{n_g, K_g\}$ which describe the self-interaction term and then hold these fixed while we optimize the remaining ones,
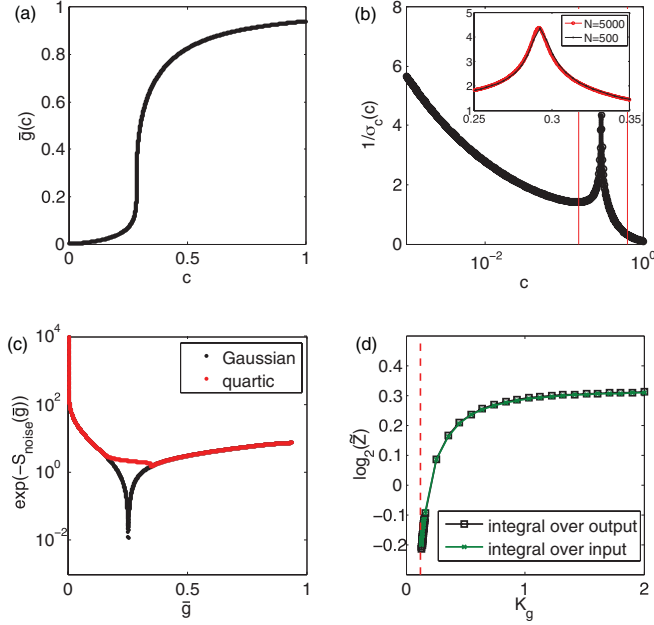
FIG. 3. (Color online) Computing information transmission close to the critical point. (a) An input-output relation $\bar{g}(c)$ for $n_g = 2$, $C = 1$, and $K_g = K_g^*(n_g) + 0.004$, showing a self-activating gene with an almost critical value of K_g. (b) Noise in the input from Eq. (14), which is the integrand in the expression for information in Eq. (36), shows an incipient divergence between red vertical bars. (Inset) Zoom-in of the peak shows that it can be sampled well by increasing the number of bins. Different plot symbols indicate domain discretization into 500 (black dots) and 5000 (red circles) bins. At the critical point the divergence is hard to control numerically. (c) An alternative way of computing the same information, by integrating in the output domain as in Eq. (69). Shown is the integrand in the Gaussian noise approximation [black, $S_2(\bar{g})$ from Eq. (66)] and with quartic corrections [red (light gray), $S_4(\bar{g})$ from Eq. (67)]. At the critical point higher-order corrections regularize the integrand, while away from the critical point the integrand smoothly joins with the Gaussian approximation. This approach is stable numerically both away from and in the critical regime. (d) Information $I = \log_2 \tilde{Z}$ as a function of K_g for $n_g = 2$. Critical $K_g^* = 1/8$ is denoted by a dashed red line. Integration across $\bar{g}$ in the output domain with quartic corrections (squares) agrees well with the integration across c in the input domain (crosses) away from K_g^*, but also smoothly extends to the critical K_g^*. This is a cut across the capacity plane in Fig. 4(a) (denoted by a dashed yellow line) for $C = 1$.

$\theta_c \equiv \{c_{1/2}, n_c, K_c\}$. In all these optimizations the parameter K_c is driven to zero, and, in this limit, the MWC regulatory function of Eq. (38) simplifies to something more like the Hill function,

$$f(g,c) = \frac{c^{n_c}}{c^{n_c} + c_{1/2}^{n_c} e^{-\tilde{F}(g)}}, \tag{88}$$

$$\tilde{F}(g) = n_g \ln\left[\frac{1 + g/K_g}{1 + 1/(2K_g)}\right]. \tag{89}$$

Once we have optimized θ_c, we can explore the information capacity as a function of θ_g, at varying values of the remaining parameter in the problem, the maximal concentration C of transcription factors. Figure 4(a) maps out the "capacity planes," $I^*(n_g, K_g; C)$ at fixed C. In detail, we show

$I^*(n_g, K_g; C) - I_{\max}^*(C)$ for three choices of our parameter C, where $I_{\max}^*(C)$ is the information obtained with the optimal choice of n_g and K_g; the best choice of parameters is depicted as a yellow circle in the capacity plane.

For large values of C, $C = 1, 10$, the optimal solution is at $n_g \to 0$ or $K_g \to \infty$ (magenta square in the lower right corner), which drives the self-activation term in Eq. (38) to zero, toward a noninteracting solution. We have checked that these solutions correspond to optimal solutions for a single noninteracting gene found in our previous work [7]. As C is decreased, however, the optimal combination $\{n_g, K_g\}$ shifts toward the left in the capacity plane (cyan square for $C = 0.1$), exhibiting a shallow but distinct maximum in information transmission. If we examine the mean input-output relations in Fig. 4(b), we find nothing dramatic: the critical (red) solutions seem to have lower capacities (which we carefully reexamine blow), while other quite distinct parameter choices for $\{n_g, K_g\}$ nevertheless generate very similar mean input-output relations, because of the freedom to optimally choose θ_c parameters. The behavior of effective noise in the input, $\sigma_c(c)$, given by Eq. (14) and shown in Fig. 4(c), is more informative; recall that $\int dc \, \sigma_c^{-1}(c)$ is proportional to the information transmission. Noninteracting (magenta) solutions always have the lowest amount of noise at high input concentrations ($c \sim C$). As the self-interaction turns on, the noise at high input increases, but that increase can be traded off for a decrease in noise at medium and low c. While for low $C = 0.1$ the critical (red) solution is never optimal, the solution with some self-activation manages to deliver an additional ~ 0.2 bits of information. We have verified that for $10\times$ smaller value of $C = 0.01$ the capacity plane is qualitatively the same, exhibiting the peak at a nontrivial (but still not critical) choice of $\{n_g \approx 0.51, K_g \approx 0.11\}$ (not shown).

Intuitively, the self-activation parameters θ_g have three direct effects on the information transmission: they change the shape of the input-output curve, the self-activation feeds some of the output noise back into the input, and the time τ (protein lifetime) that averages the input noise component gets renormalized to $\tau \to \tau(1 - \partial f/\partial g)^{-1}$. The changes in the mean input-output relation can be partially compensated for by the correlated changes in the θ_c, as we observed in our optimal solutions, suggesting that regardless of the underlying microscopic parameters, it is the shape of $\bar{g}(c)$ itself that must be optimal. The increase in averaging time acts to increase the information, thus favoring self-activation. However, this will simultaneously increase the noise in the output that feeds back, as well as drive $\bar{g}(c)$ toward infinite steepness at criticality, restricting the dynamic range of the output. At low C there is a parameter regime where increasing the integration time will help decrease the (dominant) input noise enough to result in a net gain of information. At high C, input noise is minimal and, thus, this averaging effect loses its advantage; instead, feedback simply acts to increase the total noise by reinjecting the output noise at the input, so optimizing information transmission drives the self-interaction to zero.

Next we examine in detail the behavior of information transmission close to the critical region. Close to, but not at, the critical point, we perform very fine discretization of the input range to evaluate the integral in Eq. (36), as reported in Fig. 3(b). To validate that the information indeed reaches a
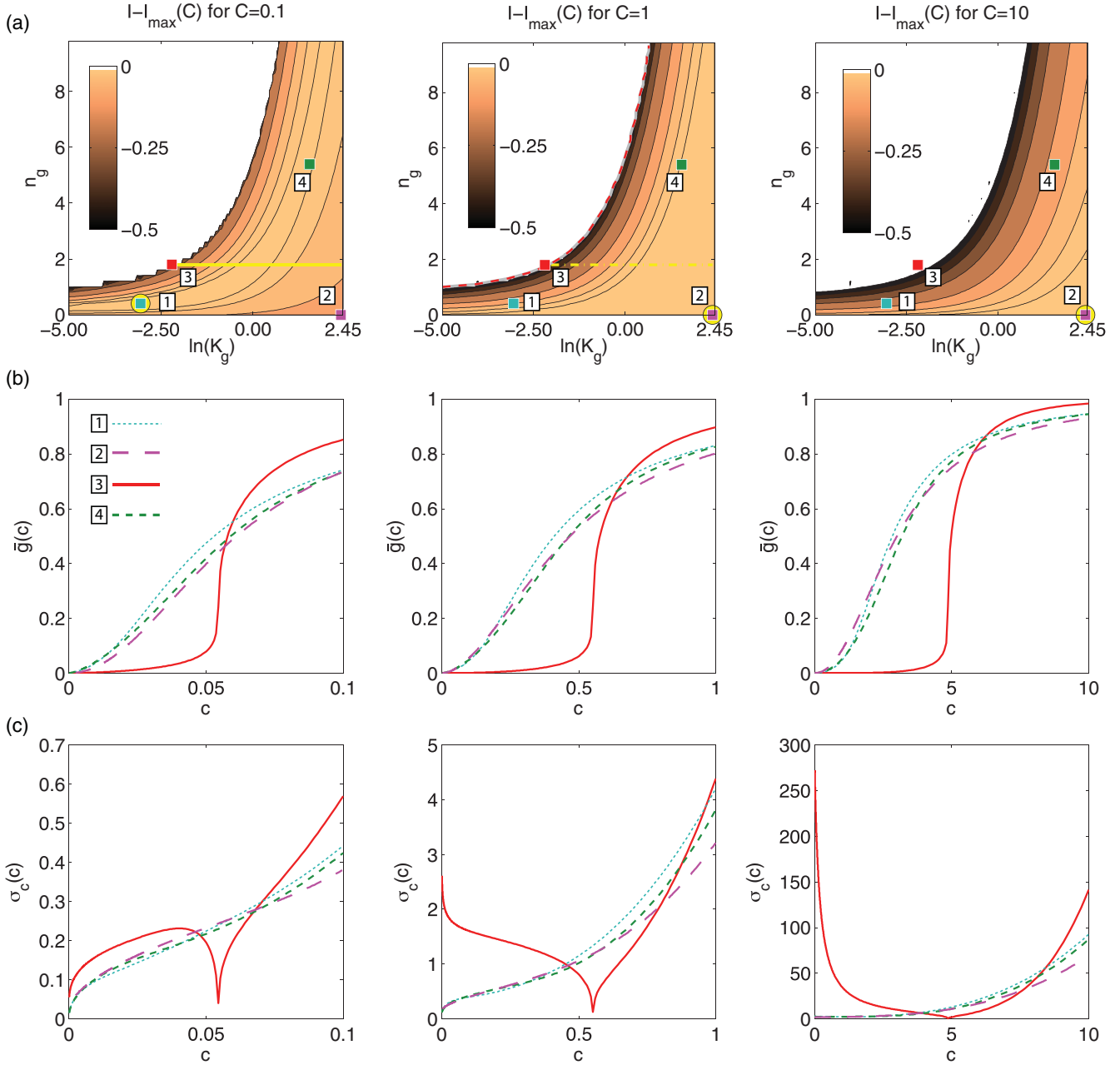
FIG. 4. (Color online) Information transmission as a function of self-activating parameters $\{n_g, K_g\}$ for three values of C; $C = 0.1$ (left column), $C = 1$ (middle column), $C = 10$ (right column). For each value of $\{n_g, K_g\}$ the three remaining parameters $\{c_{1/2}, n_c, K_c\}$ are optimized to maximize information. (a) The capacity planes showing the decrease in capacity in bits (colormap) from the maximal value achieved for an optimal choice of all parameters; the optimal set of parameters is denoted by a colored square on a yellow circle. The white upper left region of each capacity plane corresponds to the bistable region. For low $C = 0.1$ the maximum is achieved in the interior of the domain (cyan, line (1)), while for $C = 1, 10$ n_g is driven to 0 (magenta, line (2)), corresponding to the non-self-interacting solution. Red line (3) represents an example critical system at $n_g = 2$ and the green line (4) represents a self-interacting system with a high value for n_g. The yellow dashed line in $C = 1$ plane represents a cut shown in detail in Fig. 3(d). The yellow solid line in $C = 0.1$ plane represents a cut shown in detail in Fig. 5(a). The red dashed boundary of the critical region in $C = 1$ plane is analyzed in detail in Figs. 5(b) and 5(c). (b) Input-output relations for four example systems denoted by colored lines in (a). Despite substantially different values for $\{n_g, K_g\}$, the optimization of remaining parameters makes the input-output curves look very similar to the optimal solutions, except for the critical [red (solid)] curves. (c) The effective input noise $\sigma_c(c)$ for the selected systems.

maximum at nontrivial values of $\{n_g, K_g\}$ when $C = 0.1$, we cut through the capacity plane in Fig. 4(a) along the yellow line at $n_g = 2$ and display the resulting capacity values in Fig. 5(a) (the results are numerically stable when integrated on 10^4

or 10^3 points). Unlike for $C = 1$ and $C = 10$, for $C = 0.1$ the maximum is clearly achieved for a nontrivial value of K_g, but away from the critical line, confirming our previous observations. We further examine the capacity directly on
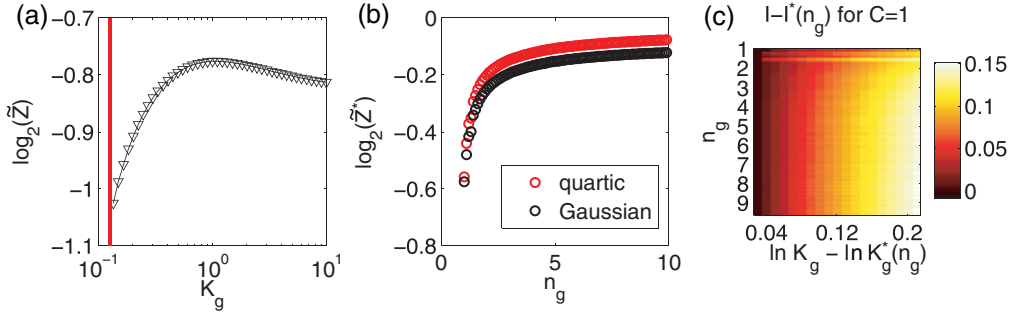
FIG. 5. (Color online) Information transmission close to the critical line. (a) A detailed scan of information for $C = 0.1$ as the function of K_g for $n_g = 2$ and optimal choice for θ_c for every K_g exhibits a peak for a nontrivial value of K_g. (b) The information transmission along the critical line for $C = 1$ as a function of n_g indicates that while including quartic corrections is important (upper line), the information at the critical line does not exhibit any large jumps (not shown). (c) A detailed scan of capacity close to the critical line for $C = 1$ as a function of n_g (vertical axis). Shown in color is the increase in information in bits compared to the value very close to the critical $I^*(n_g)$, as a function of the distance in K_g from the critical value (horizontal axis). Across the range of n_g, going from the critical axis into the monostable domain increases the information.

the critical line, $K_g^* = (n_g - 1)^2/(4n_g)$, as a function of n_g at $C = 1$ [denoted in Fig. 4(a) with a dashed red line]. The capacity in this case can be calculated using Eq. (69) and is shown in Fig. 5(b). The capacity that includes quartic corrections is higher by ~ 0.05–0.1 bits than in the Gaussian approximation, making the effect small but noticeable. We also confirmed that the capacity at the critical line joints smoothly with the capacity near the line, i.e., that there is no jump in capacity exactly at criticality, which presumably would be a sign of numerical errors. Figure 5(c) finally validates that across the whole range of n_g for $C = 1$, small increases in K_g above the critical value $K_g^*(n_g)$ always lead to an increase of information, demonstrating that the maximum is *not* achieved on the critical line.

We next turn to the joint optimization of all parameters and plot the information transmission as a function of C in Fig. 6. As we have discussed, optimization drives the strength of self-activation to zero for $C > 1$ (but see below for self-repression), and at these high values of C the result of full optimization coincides with the noninteracting case. As C falls below one, the gain in information due to self-activation is increased, reaching a significant value of about a bit for $C = 0.01$. As we have noted in Sec. II C, the self-activating effect of g on its own expression can be changed into a self-repressing effect by simply flipping the sign of the parameter n_g. To explore the optimization of such self-repressing genes, we, thus, optimized the parameters as before, now constraining $n_g \leqslant 0$. Results in $\{n_g, K_g\}$ plane are shown in Fig. 7 for $C = 1$ and $C = 10$.

We find that, for large C, the optimization process drives both K_c and K_g toward zero, so the effective input-output relation is given by

$$f(g,c) = \frac{c^{n_c}}{c^{n_c} + c_{1/2}^{n_c}(2g)^{-n_g}}, \tag{90}$$

with nonzero values of n_g being optimal. Why is self-repression optimal at large C, when self-activation is not? Self-repression suppresses noise at high concentrations of the input [red versus magenta curves Fig. 7(c) for $C = 10$] and allows the mean input-output curve to be more linear than in the noninteracting case [Fig. 7(b)], extending the dynamic

range of the response. Both these effects serve to increase information transmission.

It is remarkable that when we put together the self-activating and self-repressing solutions, we see that they join
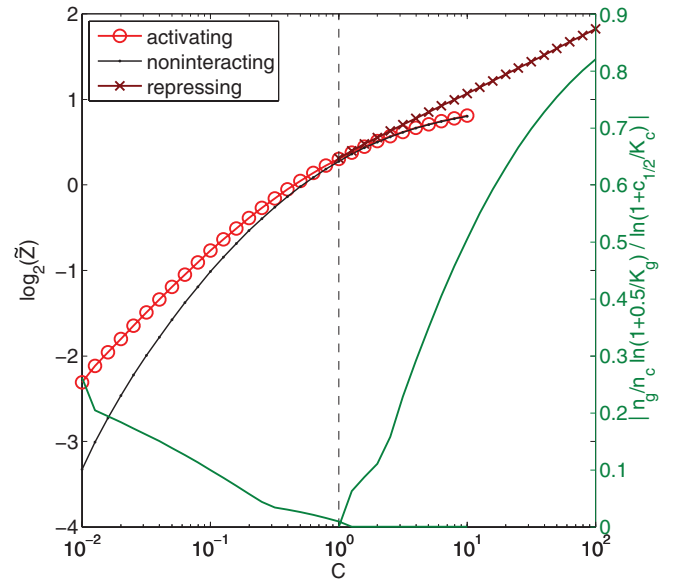

FIG. 6. (Color online) Information transmission in a system where all parameters θ are optimized as a function of the maximal input concentration C. The self-interacting system [red (crosses and circles)] allows for an arbitrary MWC-like regulatory function Eq. (38) with parameters $\theta = \{c_{1/2}, n_c, K_c, n_g, K_g\}$. The noninteracting system [black (solid dots)] only has the MWC parameters n_c, K_c and the leak L (see Ref. [7]) which can be reexpressed in terms of $c_{1/2}$. The (red) line with circles shows self-activating solutions which are optimal for $C < 1$, while the (red) line with crosses shows self-repressing solutions, optimal for $C > 1$. Plotted on the secondary vertical axis in green (solid line) is the ratio between the self-interacting contribution to F, and the input contribution to F in the expression for the MWC regulatory function Eq. (38). For $C \sim 1$ where the interacting and noninteracting solution join, this term falls to 0, as expected.
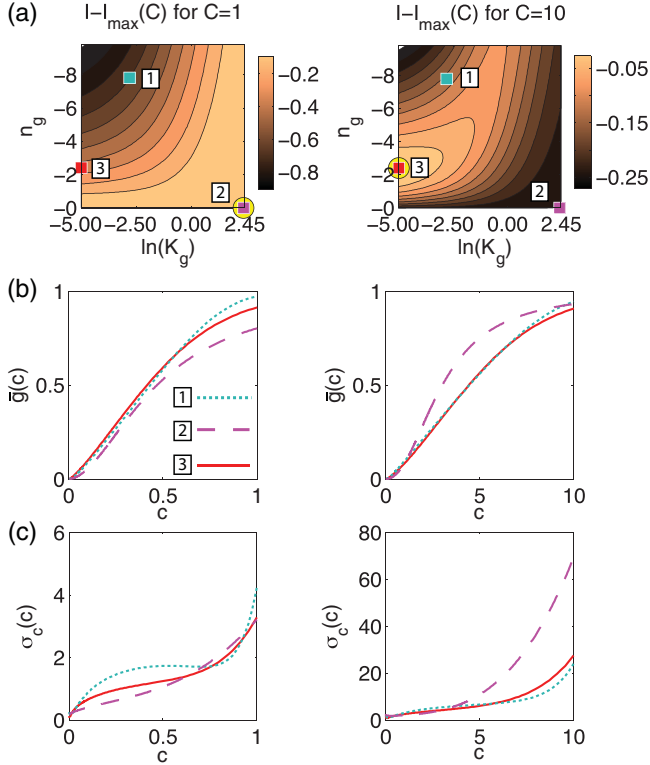
FIG. 7. (Color online) Information transmission as a function of self-repressing parameters $\{n_g, K_g\}$ for $C = 1$ (left column) and $C = 10$ (right column); plot conventions are the same as in Fig. 4. (a) The capacity decrease from the maximum value (achieved at the parameter choice indicated by a yellow circle) as a function of $\{n_g, K_g\}$. The maximum information transmission is achieved for a noninteracting case (magenta, line (2)), for $C = 1$. In contrast, for $C = 10$, there is a nontrivial optimum for small values of K_g and $n_g \approx -2.5$ (red, line (3)). (b) The mean input-output solutions for three example systems from A (cyan, line (1); magenta, line (2); red, line (3)). (c) The effective noise in the input, $\sigma_c(c)$, for example solutions in (a).

smoothly at $C = 1$ (Fig. 6): Self-activation is optimal for $C < 1$, and self-repression is optimal for $C > 1$, while precisely at $C = 1$ the system that transmits the most information is noninteracting.

All of this discussion has been in the limit where the maximal concentration of output molecules, $\gamma_{\max}$, is the same as the maximal concentration of input molecules, C, so there is only one parameter that governs the structure of the optimal solution. This makes sense, at least approximately, since both input and output molecules are transcription factors, presumably with similar costs, but, nonetheless, we would like to see what happens when we relax this assumption. Intuitively, if we let $\gamma_{\max}$ become large, the system can achieve the advantages of feedback while the impact of noise being fed back into the system should be reduced.

If we look at Eq. (36) for $\tilde{Z}$, which controls the information capacity, we can take the limit $\gamma_{\max} \rightarrow \infty$ to find

$$\tilde{Z} = \int_0^C dc \frac{\frac{\partial f}{\partial c}\left(1 - \frac{\partial f}{\partial g}\right)^{-\frac{1}{2}}}{\sqrt{\bar{g} + \left(\frac{\partial f}{\partial c}\right)^2 c}}. \qquad (91)$$
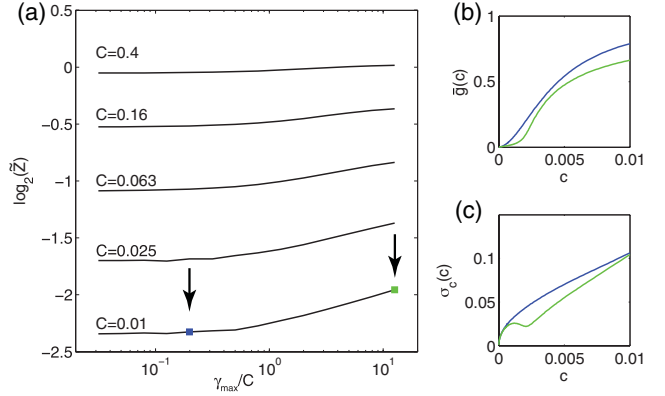

FIG. 8. (Color online) The dependence of information transmission in the self-activating case on the ratio $\gamma_{\max}/C$. (a) For various choices of C indicated on the plot, the information transmission with the optimal choice with respect to all parameters θ is shown as a function of $\gamma_{\max}/C$. Two special systems of interest [arrows, blue (dark gray) and green (light gray)] are chosen for the lowest value of C. (b) The mean input-output relation for the blue and green system. The green system has a higher transmission, a steeper activation curve but a smaller dynamic range. (c) The effective noise in the input, $\sigma_c(c)$, for the blue and green systems. The green system is closer to critical at the point where the mean input-output curve has the highest curvature and the noise exhibits a dip.

Now the only place where feedback plays an explicit role is in the term $(1 - \frac{\partial f}{\partial g})^{-\frac{1}{2}}$, which comes from the lengthening of the integration time, which, in turn, serves to average out the noise in the system. All other things being equal (which may be hard to arrange), this suggests that information transmission will be maximized if the system approaches the critical point, where $\partial f/\partial g \rightarrow 1$. The difficulty is that the system cannot stay at the critical point for all values of the input c, so there must be a trade-off between lengthening the integration time and using the full dynamic range.

To explore more quantitatively, we treat $\gamma_{\max}/C$ as a parameter. When C is small, we know that self-activation is important, and in this regime we see from Fig. 8 that changing $\gamma_{\max}/C$ matters. On the other hand, for large values of C we know that (at $\gamma_{\max}/C = 1$) optimization drives self-activation to zero, so we expect that there is less or no impact of allowing $\gamma_{\max}/C \neq 1$. We also see that, for a fixed small C, increasing $\gamma_{\max}/C$ drives the system closer toward the signatures of criticality—nonmonotonic behavior in the noise and a steepening of the input-output relation. In more detail, we can plot the value of $\max_c(\partial f/\partial g)$ as a function of C and $\gamma_{\max}/C$, that is, check for each of the solutions in Fig. 8(a) how close the partial derivative $\partial f/\partial g$ comes to 1, which is a direct measure of criticality. We confirm that, for the simultaneous choice of small C and large $\gamma_{\max}/C$, we indeed have $\partial f/\partial g \rightarrow 1$. In the extreme, if we choose $C = 0.01$ and $\gamma_{\max}/C = 10^4$, we find that the optimal K_c and K_g are driven toward small values (but since c, C are small, K_c is not negligible); the optimal $n_g \approx 1.0662$. With this value of n_g, the corresponding critical value for K_g would be $K_g^*(n_g) = 0.001$, and the numerically found optimal value in our system is $K_g = 0.0012$. The critical value for $\bar{g}^*$ would be $\bar{g}^*(n_g) = 0.031$, and, indeed, at this small value the optimal
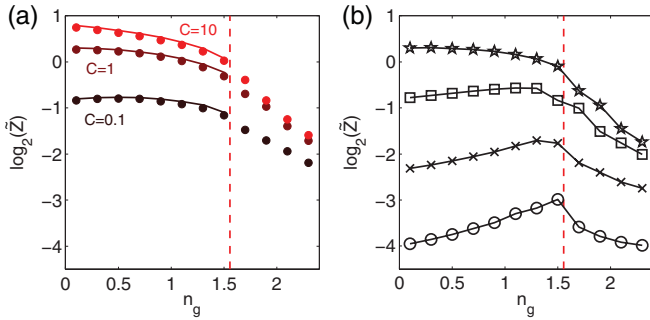
FIG. 9. (Color online) Crossing the critical line (dashed red) into the bistable region. (a) Capacity as a function of n_g at fixed $K_g = 0.05$, with the optimal choice of $\{n_c, K_c, c_{1/2}\}$ parameters, for three values of C [$C = 0.1, 1, 10$ dark (gray) to bright red (gray), respectively]. Dots show capacity calculation using the bistable code that can handle multiple branches using Eq. (85), solid line uses the monostable integration as in Eq. (36). (b) Optimal capacity at very high ratios $\gamma_{\max}/C = 10^3$ for different value of C ($C = 10^{-3}, 10^{-2}, 10^{-1}, 1$: circles, crosses, squares, and stars, respectively). The optimum is pushed toward the critical line from the monostable side for $\gamma_{\max}/C$ large and C small. In all cases, information in the bistable regime is smaller than in the monostable regime.

mean input-output relation has a strong kink, the effective noise σ_c has a sharp dip, and $\partial f/\partial_g$ at this point climbs to 0.9936. Numerically, therefore, we have all the expected indications of emerging criticality at very large $\gamma_{\max}$. For less extreme values, we expect the optimum to result from the interplay between the input and transmitted noise contributions, which, in general, need not be on the critical line.

To complete our exploration of the optimization problem, we have to consider parameter values for which the output has two locally stable values given a single input. Quantitatively, in the bistable regime we have to solve for both stable solutions $\bar{g}_i(c)$, with $i = 1, 2$, and for the unstable branch $\bar{g}_3(c)$. We can then evaluate the equilibrium probabilities $w_i(c)$ of being on either of the stable branches using Eq. (87), and use Eq. (85) to compute the capacity. As shown in an example in Fig. 9(a), we never find the optimal solutions in the bistable region–the capacity starts decreasing after crossing the critical line. Consistently with our argument that output and feedback noise must become negligible for the regime of small C and large $\gamma_{\max}/C$, we find that optimization drives the system toward achieving maximal transmission closer and closer the the critical line (which is approached from the monostable side), as shown in Fig. 9(b).

IV. DISCUSSION

To summarize, we have analyzed in detail a single, self-interacting genetic regulatory element. As in previous work, we based our analysis on three assumptions: (i) that the readout of the information $I(c; g)$ between the input and output happens in steady state, (ii) that noise is small; and (iii) that the constraint limiting the information flow is the finite number of signaling molecules. In addressing a system with feedback, assumption (ii) requires technical elaboration near the critical point, as discussed above. But (i) requires a qualitatively

new discussion for systems with feedback, because of the possibility of multistability.

Our analysis, with the steady-state assumption, shows that truly bistable systems do not maximize the information. Intuitively, this stems from the branch ambiguity: For a given input concentration c a bistable system can sit on either one of the stable branches with some probability, and this uncertainty contributes to the noise entropy, thereby reducing the transmitted information. But reaching steady state involves waiting for two very different processes. First, the system reaches a steady distribution of fluctuations in the neighborhood of each stable point, and then the populations of the two stable states equilibrate with one another. As with Brownian motion in a double-well potential (or a chemical reaction), these two processes can have time scales that are separated by an exponentially large factor.

Alternatively, the time scales of real regulatory and readout processes could be such that the system does not have the time to equilibrate between the stable branches. In that case, the history (initial condition) of the system will matter, and the final value of the output g will be determined jointly by the input c and the past state of the output, g_0. Such regulatory elements can be very useful, because they retain memory of the past and are able to integrate it with the new input; a much studied biological example is that of a toggle switch. The information measure we use here, $I(c; g)$, will not properly capture the abilities of such elements, unless we modify it to include the past state, e.g., into $I(\{c, g_0\}; g)$: Here both the input and current state together determine the output. Such computations are beyond the scope of this paper but could make precise our intuitions about switches with memory.

Multistability also allows for qualitatively new effects at higher noise levels. In our previous work we found that full information flow optimization (without assuming small noise) leads to higher capacities than a small-noise calculation for an identical system and, moreover, that as noise grows, the optimal solutions start resembling a (noisy) binary switch where only the minimum and maximum states of input are populated in the optimal $P_{\text{in}}^*(c)$ [4]. At high noise, positive (even bistable) autoregulation could stabilize these two states and make them more distinguishable. In this case, the design constraint for the genetic circuit is to use the smallest number of molecules that will prevent spontaneous flipping between the two branches on the relevant biological time scales [18]. In this limit regulatory elements can operate at high noise, with perhaps as few as tens of signaling molecules.

With these caveats in mind, our main results can be summarized as follows. Except at $C \sim 1$, the possibility of self-interaction always increases the capacity of genetic regulatory elements. For $C < 1$, the optimal strategy is self-activation, while for $C > 1$ it is self-repression, as shown in Fig. 6. Self-repression allows the system to reduce the effective noise at high input levels and straighten the input-output relation, packing more "distinguishable" signaling levels into a fixed input range. Self-activation for small C lengthens the effective integration time over which the (dominant) input noise contribution is averaged, thereby increasing information. The optimal level of self-activation is never so strong as to cause bistability, but does, for small C and large $\gamma_{\max}/C$, push the optimal system toward the critical state.

An interesting observation about the nature of the optimal solutions is that self-activation which is strong enough to enhance information transmission may, nonetheless, not result in a functional input-output relation that differs in appearance from a system without self-activation, albeit with different parameters. In such cases, information transmission is enhanced primarily by the longer integration time and reduced effective noise level. This means that there need be no dramatic signature of self-activation, so diagnosing this operating regime requires a detailed quantitative analysis. More generally, this result emphasizes that the same phenomenology can result from different parameter values or even networks with different topology—in this case, with and without feedback.

The prevailing view of self-activation has been that its utility stems from the possibility of creating a toggle (or a flip-flop) switch. This explanation, however, can be true only if self-activation is strong enough to actually push the system into the bistable regime. De Ronde and colleagues [38] have improved on this intuition and have shown, in the linear response limit, that weak self-activation will increase the signal-to-noise ratio for dynamic signals, a function that differs markedly from the switch. Here we show that in the fully nonlinear, but steady-state, treatment, monostable self-activation can be advantageous for information transmission. Furthermore, we show that there is a single control parameter, the ratio C between the output and input noise strengths, which determines whether self-activation or self-repression is optimal. Since more and more quantitative expression data are available, especially for bacteria and yeast, one could try to assess how the use of both motifs correlates with the concentrations of input and output signaling molecules.

Stepping back from the detailed results, our goal in this paper was to make progress on understanding the optimization of information flow in systems with feedback by studying the simplest example. The hope is that our results provide one building block for a theory of real genetic networks, on the hypothesis that they have been selected for maximizing information transmission. As discussed in previous work [3,6,7], a natural target for such analysis is the well-studied gap gene network in the early *Drosophila* embryo [45], although we also hope to connect with a broader range of examples. Real systems, however, often use regulatory mechanisms that are more complex than the simplest setup examined here which enables theoretical progress. One of the important open questions is, therefore, how different molecular schemes of transcriptional control impact the information flow in transcriptional regulation. We started with a phenomenologically motivated MWC regulatory function and two major noise contributions in the form of the input diffusive and output shot noise. In fact, processes such as DNA looping [35], regulation of the oligomeric state of the transcription factors [46,47], regulation of decay [48], and control through chromatin state modification or the covalent modification of other regulatory molecules all potentially affect the mean input-output relation and the noise. Nevertheless, our hope is that if these quantities can be computed for more complicated architectures, the overall information-maximization framework should remain applicable. We leave open for future work the question of generalizing our results to more complicated molecular regulatory mechanisms.

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[2] G. Tkačik and A. M. Walczak, J. Phys.: Condens. Matter **23**, 153102 (2011).

[3] G. Tkačik, C. G. Callan Jr., and W. Bialek, Proc. Natl. Acad. Sci. USA **105**, 12265 (2008).

[4] G. Tkačik, C. G. Callan Jr., and W. Bialek, Phys. Rev. E **78**, 011910 (2008).

[5] M. Ronen, R. Rosenberg, B. I. Shraiman, and U. Alon, Proc. Natl. Acad. Sci. USA **99**, 10555 (2002).

[6] G. Tkačik, A. M. Walczak, and W. Bialek, Phys. Rev. E **80**, 031920 (2009).

[7] A. M. Walczak, G. Tkačik, and W. Bialek, Phys. Rev. E **81**, 041905 (2010).

[8] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).

[9] I. M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingreham, S. Paley *et al.*, Nucl. Acids Res. **33**, D334 (2005).

[10] R. Hermsen, B. Ursem, and P. R. ten Wolde, PLoS Comput. Biol. **6**, e1000813 (2010).

[11] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Curr. Opin. Genet. Dev. **15**, 116 (2005).

[12] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, Curr. Opin. Genet. Dev. **15**, 125 (2005).

[13] N. Rosenfeld, M. B. Elowitz, and U. Alon, J. Mol. Biol. **323**, 785 (2002).

[14] A. Becksei and L. Serrano, Nature **405**, 590 (2000).

[15] A. Becksei, B. Seraphin, and L. Serrano, EMBO J. **20**, 2528 (2001).

[16] N. E. Buchler, U. Gerland, and T. Hwa, Proc. Natl. Acad. Sci. USA **100**, 5136 (2003).

[17] A. Arkin, J. Ross, and H. H. McAdams, Genet. **149**, 1633 (1998).

[18] W. Bialek, in *Advances in Neural Information Processing Systems*, edited by T. K. Leen, T. G. Dietterich, and V. Tresp (MIT Press, Cambridge, 2001), pp. 103–109.

[19] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature **403**, 339 (2000).

[20] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, Cell **130**, 153 (2007).

[21] T. Erdmann, M. Howard, and P. R. ten Wolde, Phys. Rev. Lett. **103**, 258101 (2009).

[22] W. Bialek and S. Setayeshgar, Proc. Natl. Acad. Sci. USA **102**, 10040 (2005).

[23] A. Sanchez, H. G. Garcia, D. Jones, R. Phillips, and J. Kondev, PLoS Comput. Biol. **7**, e1001100 (2011).

[24] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).

[25] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. USA **98**, 8614 (2001).

[26] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. D. Swain, Science **297**, 1183 (2002).

[27] E. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).

[28] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, Nature **422**, 633 (2003).

[29] J. M. Raser and E. K. O'Shea, Science **304**, 1811 (2004).

[30] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, Science **307**, 1962 (2005).

[31] J. M. Pedraza and A. van Oudenaarden, Science **307**, 1965 (2005).

[32] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).

[33] D. Zenklusen, D. R. Larson, and R. H. Singer, Nat. Struct. Mol. Biol. **15**, 1263 (2008).

[34] L. H. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, Nat. Genet. **43**, 554 (2011).

[35] J. M. G. Vilar and S. Leibler, J. Mol. Biol. **331**, 981 (2003).

[36] L. Saiz and J. M. G. Vilar, IET Systems Biol. **2**, 247 (2008).

[37] E. Ziv, I. Nemenman, and C. H. Wiggins, PLoS ONE **2**, e1077 (2007).

[38] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys. Rev. E **82**, 031914 (2010).

[39] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer, Berlin, 2009).

[40] G. Tkačik and W. Bialek, Phys. Rev. E **79**, 051901 (2007).

[41] G. Tkačik, T. Gregor, and W. Bialek, PLoS One **3**, e2774 (2008).

[42] A. V. Hill, J. Physiol. **40**, Suppl iv-vii (1910).

[43] J. Monod, J. Wyman, and J. P. Changeux, J. Mol. Biol. **12**, 88 (1965).

[44] L. A. Mirny, Proc. Natl. Acad. Sci. USA **107**, 22534 (2010).

[45] J. Jaeger, Cell Mol. Life Sci. **68**, 243 (2011).

[46] J. M. G. Vilar and L. Saiz, Nucl. Acids Res. **39**, 6854 (2011).

[47] F. Nicol-Benoit, A. Amon, C. Vaillant, P. le Goff, Y. le Drean, F. Pakdel, G. Flouriot, Y. Valotaire, and D. Michel, Biophys. J. **101**, 1557 (2011).

[48] N. E. Buchler, U. Gerland, and T. Hwa, Proc. Natl. Acad. Sci. USA **102**, 9559 (2005).