Optimizing information flow in small genetic networks. II. Feed-forward interactions

Aleksandra M. Walczak,[1,*] Gašper Tkačik,[2,†] and William Bialek[1,‡]
[1]*Joseph Henry Laboratories of Physics, Lewis-Sigler Institute for Integrative Genomics, and Princeton Center for Theoretical Science, Princeton University, Princeton, New Jersey 08544, USA*
[2]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6396, USA*
(Received 31 December 2009; published 6 April 2010)

Central to the functioning of a living cell is its ability to control the readout or expression of information encoded in the genome. In many cases, a single transcription factor protein activates or represses the expression of many genes. As the concentration of the transcription factor varies, the target genes thus undergo correlated changes, and this redundancy limits the ability of the cell to transmit information about input signals. We explore how interactions among the target genes can reduce this redundancy and optimize information transmission. Our discussion builds on recent work [Tkačik *et al.*, Phys. Rev. E **80**, 031920 (2009)], and there are connections to much earlier work on the role of lateral inhibition in enhancing the efficiency of information transmission in neural circuits; for simplicity we consider here the case where the interactions have a feed forward structure, with no loops. Even with this limitation, the networks that optimize information transmission have a structure reminiscent of the networks found in real biological systems.

DOI: 10.1103/PhysRevE.81.041905 PACS number(s): 87.10.Vg, 87.15.R−, 87.18.Tt, 87.85.Ng

I. INTRODUCTION

The genomes of even the smallest bacteria encode the structure of hundreds of proteins; for complicated organisms like us, the number of different proteins reaches into the tens of thousands [1]. During the course of its life, each cell has to control how many copies of each protein molecule are synthesized [2]. These decisions about the expression of genetic information occur on many scales, from (more or less) irreversible decisions during differentiation into different tissue types down to continuous modulations of the number of enzyme molecules in response to varying metabolic needs and resources [3]. As with many biological processes, the control of gene expression depends critically on the transmission of information. Because the relevant information is represented inside the cell by signaling molecules at low concentrations, the irreducibly stochastic behavior of individual molecules sets a physical limit to information transmission [4–7]. A more precise discussion is motivated by the emergence of experiments which measure the noise in gene expression [8–15]. In this paper, building on our previous work [16], we explore how cells can structure their genetic control circuitry to optimize information transmission in the presence of these physical constraints, thus maximizing the control power that they can achieve while using only a limited number of molecules.

Although the problems of information transmission in genetic control are general, it is useful to have a concrete example in mind. In developing embryos, information about position—and hence, ultimately, fate in the developed organism—is encoded by spatially varying concentrations of "morphogen" molecules [17,18]. These molecules often are transcription factors, and positional information then is transmitted to the expression levels of the target genes [18–20]. In this scheme, maximizing information transmission maximizes the richness of the spatial patterns which the embryo can construct.

In recent work we have considered the problem of maximizing information transmission from a single transcription factor to multiple target genes, in the case where the targets are noninteracting [16]. Already, this problem generates some structures that remind us of real genetic control networks. But, when a single transcription factor controls many genes independently, the signals carried by those genes necessarily are redundant [21]. To fully optimize the information which can be carried by a given number of molecules, there must be interactions among the target genes which reduce this redundancy. Our goal in this paper is to derive the form of these interactions. We emphasize that this is an ambitious project: we are trying to find the topology and parameters of a genetic network from first principles, constrained only by the limited concentration of all the relevant molecules. To simplify our task, we start here with the case in which the network of interactions has no closed loops, and return to the full problem in subsequent papers.

When we search possible networks for the ones that optimize information transmission, it is convenient to break this search into two parts. First, we enumerate network topologies, and then we search parameters within each topology. It is conventional to encode in the network topology the sign of the interactions, so that changing the 'shape of an arrow' from activation to repression is counted as a qualitatively different network (see Fig. 1). Even if we assume that each protein can act either as activator or repressor for all its targets [22], with one transcription factor controlling K interacting genes in a feed forward network, there are 2^K possible network topologies. For each of these possibilities we find a single well defined optimum for all the continuously adjustable parameters in the system, and these optimal parameter values are determined only by the number of available molecules. Thus, the structures of these networks really are derivable, quantitatively, from first principles. For related ap-

*awalczak@princeton.edu
†gtkacik@sas.upenn.edu
‡wbialek@princeton.edu

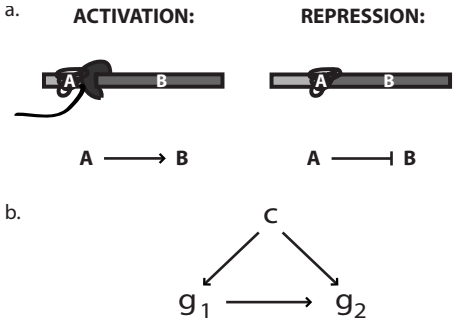

FIG. 1. Schematic models of transcriptional regulation. (a) Transcription factor proteins (A) can either activate (left) the expression of genes (B) by recruiting the RNA polymerase, or repress (right) the expression of genes by blocking the polymerase. Throughout the paper we will use the diagrammatic representation, in which an arrow depicts activation and a blunted arrow depicts repression. (b) A feed forward network in which the input c regulates the expression levels of g_1 and g_2, and g_1 also regulates g_2; the structure is feed forward because g_2 does not feedback on g_1.

proaches to information flow in biochemical and genetic networks, see Refs. [24–28].

Among the possible network topologies, there are some in which interaction strengths are driven to zero by the optimization. Among the nontrivial solutions, however, we find that the different topologies achieve very similar information capacities. Thus, it seems that there is not a single optimal network, but rather an exponentially large number of nearly degenerate local optima. All of these local optima are networks with competing interactions, so that a single target gene is activated and repressed by different inputs. Phenomenologically, if we view the expression level of each target gene as a function of the input transcription factor concentration, these competing interactions lead to nonmonotonic input/output relations. It is this nonmonotonicity that allows the collection of target genes to explore more fully the space of expression levels, thus enhancing the capacity to transmit information. Although we consider only networks without feedback, these nonmonotonic responses already are reminiscent of the patterns seen in real genetic networks.

II. SETTING UP THE PROBLEM

The optimization problem that we are addressing in this paper is a generalization of the problem discussed in our previous work [16]. To make this paper self-contained, however, we begin by reviewing the general formulation of the problem. We are interested in situations where a single transcription factor controls the expression level of several genes, labeled $i=1,2,\ldots,K$. The expression levels of these genes, g_i, carry information about the concentration c of the input transcription factor. It is this information $I(c;\{g_i\})$ that we suggest is optimized by real networks.

To derive the consequences of our optimization principle, we require several ingredients: we need to describe the space of possible networks, we need to relate $I(c;\{g_i\})$ to some calculable properties of these networks, we need to put these pieces together to form the objective function for our optimization problem, and finally we need to perform the optimization itself. In the interests of proceeding analytically as far as possible down this path, before resorting to numerics, we adopt several approximations: we consider the case where expression levels have reached steady state values consistent with the input transcription factor concentration, we assume that the noise in the system is small, and we restrict our attention to networks that have a feed forward structure. We have discussed the first two approximations in our previous work [16], and our focus here on feed forward networks is intended as a useful intermediate step between the case of noninteracting target genes (as in Ref. [16]) and the case of arbitrary interactions, to which we will return in a subsequent paper.

A. Describing the regulatory interactions

The expression levels of genes are determined by a complex sequence of events, especially in eukaryotic cells where even the transcription of DNA into messenger RNA involves a complex of more than one hundred proteins [23]. As in much previous work, we abstract from this complexity to say that expression levels are set by a balance of synthesis and degradation reactions; in the simplest case only synthesis is regulated, and degradation is a first order process. Then the deterministic kinetic equations that govern the expression levels are

$$\frac{dG_i}{dt} = r_{\max} f_i(c;\{G_j\}) - \frac{1}{\tau} G_i, \tag{1}$$

where τ is the lifetime of the gene products against degradation, $r_{\max}$ is the maximum number of molecules per second that can be synthesized, and $f_i(c;\{G_j\})$ is the "regulation function" that describes how the synthesis rate is modulated by the other molecules in system. The regulation functions, which are positive and range between zero and one, will be different for every gene, but we assume that the lifetimes and maximum synthesis rates are the same for all of the genes [29]. In this formulation, G_i counts the number of molecules of the protein coded by gene i, while it is conventional to define c as a concentration. Before we are done, it will be useful to normalize these quantities.

The steady state expression levels, $\bar{G}_i(c)$, are the solutions of the simultaneous nonlinear equations

$$\bar{G}_i(c) = r_{\max} \tau f_i(c;\{\bar{G}_j(c)\}). \tag{2}$$

In general there can be multiple solutions, corresponding to a network that has more than one stationary state; as noted above, we will confine our attention here to networks with a feed forward architecture, in which this cannot happen. In Fig. 1(b), we present an example of a feed forward network, where an input c regulates the expression of genes g_1 and g_2. The product proteins of g_1 also regulate gene g_2, but g_2 does not regulate g_1. Formally, the "no loops" condition defining feed forward networks means that there is some assignment of the labels i so that f_i only depends upon G_j for $j<i$, and hence, as will be important below, the matrix $\partial f_i/\partial G_j$ is lower triangular.

A deterministic system with continuous inputs and outputs can transmit an infinite amount of information. In reality, information transmission is limited by the discreteness and randomness of the individual molecular events. If this noise is small, it can be described by adding Langevin "forces" to the deterministic kinetic equations [30–34]. To the extent that synthesis and degradation rates can be written as depending on the instantaneous concentrations or expression levels, as we have done in Eq. (1), then the system has no "hidden" memory, and the Langevin terms which describe the noise in the system should have effectively zero correlation time. We also will assume that the noise in the synthesis and degradation of different gene products are independent. Then we can write

$$\frac{dG_i}{dt} = r_{\max} f_i(c;\{G_j\}) - \frac{1}{\tau} G_i + \eta_i(t), \tag{3}$$

$$\langle \eta_i(t)\, \eta_j(t') \rangle = \mathcal{N}_i \delta_{ij} \delta(t-t'), \tag{4}$$

where the $\mathcal{N}_i$ are the spectral densities of the noise forces.

The spectral densities of noise have several components which add together. First, if we take Eq. (1) seriously, it describes synthesis and degradation as simple first order processes; the terms which appear in the deterministic equation as rates should then be interpreted as the mean rates of Poisson processes that describe the individual molecular transitions. Then we have

$$\mathcal{N}_i^{(1)} = r_{\max} f_i(c;\{G_j\}) + \frac{1}{\tau} G_i. \tag{5}$$

We are interested in small fluctuations around the steady state, so in this expression we should set $G_i = \bar{G}_i(c)$. Further, we can account for the possibility that synthesis is a multistep process by adding a "Fano factor" ν to the spectral density of synthesis noise, so we have

$$\mathcal{N}_i^{(1)} = \nu r_{\max} f_i(c;\{\bar{G}_j(c)\}) + \frac{1}{\tau} \bar{G}_i(c), \tag{6}$$

$$\mathcal{N}_i^{(1)} = \frac{1+\nu}{\tau} \bar{G}_i(c). \tag{7}$$

The Fano factor is a measure of the dispersion of the probability distribution, the ratio of the variance to the mean; here the relevant random variable is the number of reaction events in a small window of time. As an example, if synthesis occurs in bursts, then we expect $\nu > 1$ [15,31,35,36]. Bursts of both proteins and mRNAs have now been experimentally observed in a number of bacterial and eukaryotic systems [9,37–39].

A second source of noise comes from the diffusive arrival of transcription factor molecules at their targets. To describe this, we should remember that the regulation of transcription depends on events that occur in a very small volume, of linear dimension ℓ. The relevant concentrations are those in this volume, or more colloquially at the point where the transcription factors bind; binding and unbinding events act as localized sources and sinks for the diffusion equation that describes the spatiotemporal variations in concentration. Intuitively, this coupling generates noise because the diffusion of the transcription factors into the relevant volume is a stochastic process, so that even with the global concentration fixed there are local concentration fluctuations [40]. Analysis of this problem [4,5,15,41] shows that, so long as the time for diffusion through a distance ℓ is short, the net effect of the diffusive fluctuations is captured by a Langevin term

$$\mathcal{N}_i^{(2)} = \left(r_{\max} \frac{\partial f_i(c;\{G_j\})}{\partial c} \right)^2 \frac{c}{D\ell}, \tag{8}$$

where D is the diffusion constant; again we should evaluate this at $G_j = \bar{G}_j(c)$ to describe the small fluctuations around the steady state.

The same arguments that apply to the input transcription factor also apply to each of the target gene products when they act as transcription factors. If there are G_i molecules of gene product i, then these proteins are present at concentration G_i/Ω, where Ω is the relevant volume [42], and the analog of Eq. (8) is

$$\mathcal{N}_i^{(3)} = \sum_{\mathrm{k}} \left(r_{\max} \frac{\partial f_i(c;\{G_j\})}{\partial (G_{\mathrm{k}}/\Omega)} \right)^2 \frac{(G_{\mathrm{k}}/\Omega)}{D\ell}, \tag{9}$$

$$= \frac{\Omega}{D\ell} \sum_{\mathrm{k}} \left(r_{\max} \frac{\partial f_i(c;\{G_j\})}{\partial G_{\mathrm{k}}} \right)^2 G_{\mathrm{k}}. \tag{10}$$

Finally, in putting all these terms together, it is convenient to measure expression levels in units of the maximum mean expression level, which is $\bar{G}_i = r_{\max}\tau$ molecules; that is, we define $g_i = G_i/(r_{\max}\tau)$. Then we have

$$\tau \frac{dg_i}{dt} = f_i(c;\{g_j\}) - g_i + \xi_i(t), \tag{11}$$

$$\langle \xi_i(t) \xi_j(t') \rangle = \delta_{ij} \delta(t-t') \frac{1}{r_{\max}^2} [\mathcal{N}_i^{(1)} + \mathcal{N}_i^{(2)} + \mathcal{N}_i^{(3)}], \tag{12}$$

$$= \delta_{ij} \delta(t-t') \tau \cdot \left[\frac{1+\nu}{r_{\max}\tau} \bar{g}_i(c) + \left(\frac{\partial f_i(c;\{g_\ell\})}{\partial c} \right)^2 \frac{c}{D\ell\tau} + \sum_{\mathrm{k}} \left(\frac{\partial f_i(c;\{g_\ell\})}{\partial g_{\mathrm{k}}} \right)^2 \frac{g_{\mathrm{k}}}{D\ell\tau(r_{\max}\tau/\Omega)} \right] \Bigg|_{\{g_{\mathrm{k}}=\bar{g}_{\mathrm{k}}(c)\}}. \tag{13}$$

The parameter combinations which appear in these equations have simple interpretations. If $\nu = 1$, the synthesis and degradation of each gene product molecule really is an independent Poisson process, and hence $r_{\max}\tau$ is not just the maximum number of molecules that can be made (on average), but also the maximum number of *independent* molecules, N_g in the notation of Ref. [16]. For more complex processes, where $\nu \neq 1$, we can write $N_g = 2r_{\max}\tau/(1+\nu)$. Similarly, the combination $r_{\max}\tau/\Omega$ is a concentration, the maximum (mean) concentration of the gene products. Since these are acting as transcription factors, we will assume that this maximum concentration is also the maximum concentra-

tion of the input transcription factor, c_{max}. As discussed previously [16], there is a natural scale of concentration, $c_0 = N_g/(D\ell\tau)$, and if we use units in which $c_0 = 1$ we can write

$$\langle \xi_i(t)\xi_j(t')\rangle = \delta(t-t')N_{ij} = \delta(t-t')\delta_{ij}\frac{\tau}{N_g}\left[\bar{g}_i(c) + \left(\frac{\partial f_i(c;\{g_\ell\})}{\partial c}\right)^2 c + \frac{1}{c_{max}}\sum_k \left(\frac{\partial f_i(c;\{g_\ell\})}{\partial g_k}\right)^2 g_k\right]\Bigg|_{\{g_k=\bar{g}_k(c)\}}. \quad (14)$$

To complete our description of the system we need to specify the regulation functions $f_i(c;\{g_j\})$.

B. Regulation functions

The central event in transcriptional regulation is the binding of the transcription factor(s) to their specific sites along the DNA. These binding sites must be close enough to the start of the coding sequence of the target gene that binding can influence the initiation of transcription by the RNA polymerase and its associated proteins. In the simplest geometric picture, which might actually be accurate in bacteria [43–46], we can imagine the molecular interactions being so strong that the normalized mean rate of transcription—what we write as the regulation function $f_i(c;\{g_j\})$—is essentially the probability that certain binding sites are occupied or unoccupied. If n transcription factors bind cooperatively to nearby sites, and binding activates transcription, then we would expect

$$f(c) = \frac{c^n}{K^n + c^n}, \quad (15)$$

and if the binding represses transcription, so that the mean rate is proportional to the fraction of empty sites, we should have

$$f(c) = \frac{K^n}{K^n + c^n}. \quad (16)$$

In each case, the constant K measures the concentration for half–occupancy of the binding sites, and so $F = -k_B T \ln K$ can be interpreted as a binding energy for the transcription factor to its specific site along the genome. In the biological literature, these models are called Hill functions, after Hill's classical discussion of oxygen binding to hemoglobin [47]. While it is convenient to distinguish between activators and repressors, we note that, within the Hill function model, we can pass smoothly from one to the other by allowing n to change sign.

The Hill function model identifies the transcription of a gene as a logical function of binding site occupancy, so that the mean rate of transcription will be proportional to the probability of the sites being occupied (or, in the case of repressors, not occupied). Thus there is a natural generalization when multiple transcription factors converge to control a single gene, as in Fig. 2: the rate of transcription should be proportional to logical functions on the occupancy of multiple binding sites. Thus, if the input transcription factor and the gene j converge to activate gene i, transcription could depend on the AND of the two binding events; if these are independent, then the regulation function becomes

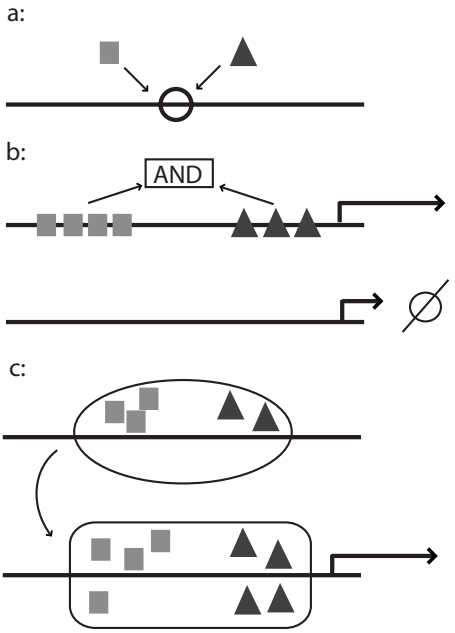

FIG. 2. Regulation of transcription in the presence of two transcription factors. (a) Two types of transcription factors, squares and triangles, can bind to the regulatory region of the gene. This situation does not uniquely determine the form of regulation. (b) The Hill regulatory model with the AND logic for the binding of the two types of transcription factors. In this case both types of transcription factors must be bound to initiate transcription. The response of the gene is proportional to the product of the activation by each transcription factor separately. (c) The MWC model for regulation. The transcriptional complex has two states, active and inactive, and binding of any type of transcription factor shifts the equilibrium between the two states.

$$f_i(c,g_j) = \frac{c^{n_i}}{K_i^{n_i} + c^{n_i}}\frac{g_j^{n_{ij}}}{K_{ij}^{n_{ij}} + g_j^{n_{ij}}}. \quad (17)$$

Alternatively, transcription could depend on the OR of the two binding site occupancies, in which case

$$f_i(c,g_j) = \frac{c^{n_i}}{K_i^{n_i} + c^{n_i}} + \frac{g_j^{n_{ij}}}{K_{ij}^{n_{ij}} + g_j^{n_{ij}}}, \quad (18)$$

and we can construct functions that interpolate or mix AND and OR. In general, we find that the pure AND functions transmit the most information. Thus, we will write

$$f_i(c,\{g_j\}) = \frac{c^{n_i}}{K_i^{n_i} + c^{n_i}}\prod_{j<i}\frac{g_j^{n_{ij}}}{K_{ij}^{n_{ij}} + g_j^{n_{ij}}}, \quad (19)$$

where the restriction $j<i$ confines our attention to feed forward networks with no loops. If $n_{ij}>0$, then $K_{ij}\to 0$ means that gene j does not regulate gene i; if $n_{ij}<0$ then the condition for no interaction is $K_{ij}\to\infty$.

An alternative model for regulation envisions a large transcriptional complex that can be in one of two states, and the rate of transcription is proportional to the population of the "active" state; transcription factors act by binding and shifting the equilibrium between the two states. This is essentially the model proposed by Monod, Wyman, and Changeaux for allosteric enzymes and cooperative binding [48]. The important idea is that, in each state, all effector molecules bind independently, but the binding energies are different in the two states. By detailed balance, this difference in binding energy means that binding will shift the equilibrium between the two states. In the case of one transcription factor at concentration c, with n binding sites, the probability of being in the active state is given by

$$f(c) = \frac{(1 + c/K_{\rm on})^n}{L^n(1 + c/K_{\rm off})^n + (1 + c/K_{\rm on})^n}, \tag{20}$$

where $k_BT \ln L$ is the free energy difference between the two states in the absence of any bound molecules. If binding favors the active state, and the gene has a very small probability to be "on" in the absence of input ($L \gg 1$), then taking $K_{\rm on} \ll K_{\rm off}$ reduces Eq. (20) to

$$f(c) = \frac{(1 + c/K_{\rm on})^n}{L^n + (1 + c/K_{\rm on})^n} \tag{21}$$

$$= \frac{1}{1 + \exp\left[- n \ln\left(\frac{1 + c/K_{\rm on}}{1 + c_{1/2}/K_{\rm on}} \right) \right]}, \tag{22}$$

where we have introduced the constant $c_{1/2}$ to simplify notation, $L=(1+c_{1/2}/K_{\rm on})$, and $c_{1/2}$ is the input concentration for half maximal activation [49]. Note that activation and repression differ only in the sign of n.

When more than one type of protein regulates expression [Fig. 2(c)], each binding event makes its independent contribution to shifting the equilibrium between active and inactive states. Thus, in the same limit of very unequal binding constants to the two states, we can write

$$f_i(c,\{g_j\}) = \frac{1}{1 + \exp[- F_i(c,\{g_j\})]}, \tag{23}$$

$$F_i(c,\{g_j\}) = n_i \ln\left(\frac{1 + c/K_i}{1 + c^{(i)}_{1/2}/K_i} \right) + \sum_{j<i} n_{ij} \ln\left(\frac{1 + g_j/K_{ij}}{1 + \bar{g}_j(c^{(i)}_{1/2})/K_{ij}} \right). \tag{24}$$

For a single transcription factor, there is not much difference between the Hill and MWC models: the classes of functions are very similar, and when we go through the optimization of information transmission, the optimal parameter values are such that the input/output relations are nearly indistinguishable. But, as we shall see, when multiple transcription factors converge, the Hill and MWC models lead to significantly different behaviors. This raises the possibility that the molecular details of the regulatory process are still 'felt' at the more macroscopic, phenomenological level, but this is a huge set of questions that we leave aside for the moment, focusing just on these two models.

C. Information transmission

We are interested in connecting the properties of the networks described in the previous section to the information that the expression levels carry about the input transcription factor concentration. This requires only a modest generalization of the discussion in Ref. [16], but, in the interest of clarity, we review the derivation.

In the limit that noise is small, the solution of the Langevin equations in Sec. II A has the form of Gaussian fluctuations around the steady state. Formally, the probability distribution of expression levels given the concentration of the input transcription factor is

$$P(\{g_i\}|c) = \frac{1}{(2\pi)^{K/2}} \exp\left\{ \frac{1}{2} \log\det(\mathcal{K}) - \frac{1}{2} \sum_{i,j=1}^{K} [g_i - \bar{g}_i(c)] \mathcal{K}_{ij} [g_j - \bar{g}_j(c)] \right\}. \tag{25}$$

The mean expression level of each gene, $\bar{g}_i(c)$, defines an input/output relation for that gene, and the matrix $\mathcal{K}$ is the inverse covariance matrix of the fluctuations or noise in the expression levels at fixed input,

$$\langle [g_i - \bar{g}_i(c)][g_j - \bar{g}_j(c)] \rangle = (\mathcal{K}^{-1})_{ij}; \tag{26}$$

note that $\mathcal{K}$ is a function of c. The fact that the expression levels "carry information" about the input transcription factor concentration means that it should be possible to estimate c from knowledge of the $\{g_i\}$. From Bayes' rule, the distribution of concentrations consistent with some set of expression levels is given by

$$P(c|\{g_i\}) = \frac{P(\{g_i\}|c)P(c)}{P(\{g_i\})}. \tag{27}$$

In the limit that noise is small ($\mathcal{K}$ is large), this too is an approximately Gaussian distribution,

$$P(c|\{g_i\}) \approx \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left[- \frac{(c - c_*(\{g_i\}))^2}{2\sigma_c^2} \right], \tag{28}$$

where $c_*(\{g_i\})$ is the most likely input given the output, and the effective variance is determined by

$$\frac{1}{\sigma_c^2(\{g_i\})} = \sum_{i,j=1}^{K} \left[\frac{d\bar{g}_i(c)}{dc} \mathcal{K}_{ij} \frac{d\bar{g}_j(c)}{dc} \right] \Bigg|_{c=c^*(\{g_i\})}. \tag{29}$$

In general it is difficult to find an explicit expression for $c_*(\{g_i\})$, but we will see that this is not crucial. We can use these ingredients to calculate the amount of information, in bits, that the expression levels $\{g_i\}$ provide about the input concentration c.

The mutual information between c and $\{g_i\}$ is defined, following Shannon [50,51], as

$$I(c;\{g_i\}) = \int dc \int d^K g P(c,\{g_i\}) \log_2 \left[\frac{P(c,\{g_i\})}{P(c)P(\{g_i\})} \right]. \tag{30}$$

We can rewrite this as a difference in entropies,

$$I(c;\{g_i\}) = -\int dc P(c) \log_2 P(c) - \int d^K g P(\{g_i\}) \times \left[-\int dc P(c|\{g_i\}) \log_2 P(c|\{g_i\}) \right], \tag{31}$$

where the first term is the entropy of the overall distribution of input concentrations, and the second term is the (mean) entropy of the distribution of inputs conditional on the output. With the Gaussian approximation from Eq. (28), we can evaluate

$$-\int dc P(c|\{g_i\}) \log_2 P(c|\{g_i\}) = \frac{1}{2} \log_2 [2\pi e \sigma_c^2(\{g_i\})]. \tag{32}$$

Further, when the noise is small, we can replace an average over expression levels by an average over inputs, setting the expression levels to their mean values along this path:

$$\int d^K g P(\{g_i\}) [\cdots] \approx \int dc P(c) \prod_i \delta(g_i - \bar{g}_i(c)) [\cdots]. \tag{33}$$

Putting these pieces together, we find an expression for the information in the low noise limit,

$$I(c;\{g_i\}) = -\int dc P(c) \log_2 P(c) - \frac{1}{2} \int dc P(c) \log_2 [2\pi e \sigma_c^2(\{\bar{g}_i(c)\})]. \tag{34}$$

As Eq. (34) makes clear, the information transmitted through the regulatory network depends on the input/output relations, on the noise level, *and* on the distribution of inputs $P(c)$. The cell can optimize information transmission by adjusting this input distribution to match the characteristics of the network [6,7], but this matching is constrained by the cost of making the input molecules. We can implement this constraint by trading bits against a cost function that counts the mean number of molecules, or more simply by insisting that c is bounded by some maximum concentration $c_{\max}$. In Ref. [16], we have shown that these different ways of implementing the constraint give essentially identical results, so we use the $c_{\max}$ approach here. We still need a constraint to force the normalization of $P(c)$, so we should maximize the functional

$$\mathcal{L} = I(c,\{g_i\}) - \lambda \int dc P(c). \tag{35}$$

The solution to this problem can be written, using the expressions for $I(c,\{g_i\})$ from above, as

$$P^*(c) \propto \frac{1}{\sigma_c(\{\bar{g}_i(c)\})} \tag{36}$$

$$= \frac{1}{Z} \left[\frac{1}{2\pi e} \sum_{i,j=1}^{K} \frac{d\bar{g}_i(c)}{dc} \mathcal{K}_{ij}(c) \frac{d\bar{g}_j(c)}{dc} \right]^{1/2}, \tag{37}$$

where the normalization constant

$$Z = \int_0^{c_{\max}} dc \left[\frac{1}{2\pi e} \sum_{i,j=1}^{K} \frac{d\bar{g}_i(c)}{dc} \mathcal{K}_{ij}(c) \frac{d\bar{g}_j(c)}{dc} \right]^{1/2}. \tag{38}$$

Finally, the information itself, evaluated with the optimal input distribution, is

$$I^*(c;\{g_i\}) = \log_2 Z. \tag{39}$$

This maximal mutual information still depends on all the parameters that describe the network, and our goal is to find the parameter values that maximize I^*. To do this, it is useful to be a little more explicit about how the parameters enter into the computation of Z.

D. Putting the pieces together

To evaluate Z in Eq. (38), we need to know the inverse covariance matrix K_{ij} that describes the noise in gene expression. This can be calculated from the Langevin equations in Sec. II A. To do this, we recall that we are looking at the small noise limit, so we linearize the Langevin equations around the steady state $g_i = \bar{g}_i(c)$, and then Fourier transform. With $g_i(t) = \bar{g}_i + \delta g_i(t)$, and

$$\delta g_i(t) = \int \frac{d\omega}{2\pi} e^{-i\omega t} \delta\tilde{g}_i(\omega), \tag{40}$$

Equation (11) becomes

$$\begin{bmatrix} 1-i\omega\tau & 0 & 0 & .. \\ -\phi_{21} & 1-i\omega\tau & 0 & .. \\ -\phi_{31} & -\phi_{32} & 1-i\omega\tau & .. \\ .. & .. & .. & .. \end{bmatrix} \begin{bmatrix} \delta\tilde{g}_1(\omega) \\ \delta\tilde{g}_2(\omega) \\ \delta\tilde{g}_3(\omega) \\ .. \end{bmatrix} = \begin{bmatrix} \tilde{\xi}_1(\omega) \\ \tilde{\xi}_2(\omega) \\ \tilde{\xi}_3(\omega) \\ .. \end{bmatrix}, \tag{41}$$

where

$$\phi_{ij} = \left. \frac{\partial f_i(c,\{g_k\})}{\partial g_j} \right|_{\{g_k = \bar{g}_k(c)\}}. \tag{42}$$

In this expression we explicitly restrict ourselves to networks with a feed-forward architecture, which (as noted above) means that we can assign labels i so that gene j regulates gene i only if $j < i$. We recall from Eq. (14) that

$$\langle \xi_i(t) \xi_j(t') \rangle = \delta(t-t') N_{ij}, \tag{43}$$

where the noise matrix is diagonal,

$$N_{ij} = \delta_{ij}\frac{\tau}{N_g}\left[\bar{g}_i(c) + \left(\frac{\partial f_i(c;\{g_\ell\})}{\partial c}\right)^2 c + \frac{1}{c_{\max}}\sum_{\rm k}\left(\frac{\partial f_i(c;\{g_\ell\})}{\partial g_{\rm k}}\right)^2 g_{\rm k}\right]\Bigg|_{\{g_{\rm k}=\bar{g}_{\rm k}(c)\}}. \quad (44)$$

The delta function in time means that the noise is white,

$$\langle \tilde{\xi}_i(\omega)\tilde{\xi}_j^*(\omega')\rangle = 2\pi\delta(\omega-\omega')N_{ij}. \quad (45)$$

Equations (41) are of the form

$$\hat{A}(\omega)\cdot\delta\tilde{\mathbf{g}}(\omega) = \tilde{\xi}(\omega), \quad (46)$$

where $\hat{A}(\omega)$ is the matrix describing the linearized dynamics, $\delta\tilde{\mathbf{g}}(\omega)=\{\delta\tilde{\mathbf{g}}_\mathbf{1}(\omega),\delta\tilde{\mathbf{g}}_\mathbf{2}(\omega),\cdots\}$, and similarly for $\tilde{\xi}(\omega)$. Evidently, $\delta\tilde{\mathbf{g}}(\omega)=\hat{A}^{-1}(\omega)\tilde{\xi}(\omega)$. We are interested in calculating the covariance matrix of the fluctuations in $\delta\mathbf{g}$. Since these are stationary, in the frequency domain we will have

$$\langle \delta\tilde{g}_i(\omega)\delta\tilde{g}_j^*(\omega')\rangle = 2\pi\delta(\omega-\omega')S_{ij}(\omega), \quad (47)$$

and the equal time correlations that we want to calculate are related to the power spectrum through

$$\langle \delta g_i \delta g_j\rangle = \int \frac{d\omega}{2\pi}S_{ij}(\omega). \quad (48)$$

Following through the algebra, we have

$$\langle \delta\tilde{g}_i(\omega)\delta\tilde{g}_j^*(\omega')\rangle = [\hat{A}^{-1}(\omega)]_{\rm ik}[\hat{A}^{-1}(\omega')]^*_{\rm jm}\langle\tilde{\eta}_{\rm k}(\omega)\tilde{\eta}^*_{\rm m}(\omega')\rangle, \quad (49)$$

$$= [\hat{A}^{-1}(\omega)]_{\rm ik}[\hat{A}^{-1}(\omega')]^*_{\rm jm}N_{\rm km}2\pi\delta(\omega-\omega'), \quad (50)$$

$$\Rightarrow S_{ij}(\omega) = [\hat{A}^{-1}(\omega)]_{\rm ik}N_{\rm km}[\hat{A}^{-1}(\omega)]^*_{\rm jm} = [\hat{A}^{-1}(\omega)\hat{N}(\hat{A}^{-1}(\omega))^\dagger]_{ij}, \quad (51)$$

where $M^\dagger$ denotes the Hermitian conjugate of the matrix M. The covariance matrix, which is the inverse of $\mathcal{K}_{ij}$ in Eq. (26), then takes the form

$$(\mathcal{K}^{-1})_{ij} = \int \frac{d\omega}{2\pi}\{\hat{A}^{-1}(\omega)\hat{N}[\hat{A}^{-1}(\omega)]^\dagger\}_{ij}. \quad (52)$$

To complete the calculation of Z we need to do this frequency integral, and then an integral over the concentration c, as in Eq. (38). Some details of these calculations are given in the Appendix. Here we draw attention to two key points.

First, because the noise ξ is white—that is, $\hat{N}$ is independent of frequency—all of the structure in the frequency integral comes from the inverse of the matrix $\hat{A}(\omega)$. Since the matrix elements of $\hat{A}$ are linear in frequency, this results in an integrand which is a rational function of ω, so the integrals can all be done by closing a contour in the complex ω plane. Again, examples are in the Appendix.

Second, and most importantly, when we finish computing Z it depends on all the parameters we want to optimize—e.g., in the Hill description of the regulation functions, all of the half-maximal points $\mathcal{K}_{ij}$ and the cooperativities n_{ij}—and in addition it depends on the maximal concentration of the input transcription factor, $c_{\max}$. But there are no other parameters in the problem. Thus, if we search for the maximum of Z, and hence the optimum information transmission, we will find the parameters of the network as a function of the available number of molecules, with no remaining arbitrariness.

III. OPTIMAL NETWORKS AND THEIR PARAMETERS

In previous work [16], we considered the case of many noninteracting genes regulated by a single transcription factor. Optimizing information transmission in these systems leads to two qualitatively different families of solutions, depending on the available dynamic range $c_{\max}/c_0$. At small $c_{\max}$, the optimal strategy is for the different output genes to respond in the same way, allowing the system to make multiple independent "measurements" of the input concentration; in this case the target genes are completely redundant. At large $c_{\max}$, a tiling solution appears, in which the genes are turned on sequentially at different concentrations; the precise setting of the thresholds is determined by a compromise between minimizing the noise and maximizing the use of the dynamic range. Even in this regime, however, the target genes are (partially) redundant. Here we explore how the interactions between target genes can be tuned to reduce redundancy and increase the capacity of the system to transmit information.

A. Two target genes

We first consider feed forward networks in which an input transcription factor at concentration c regulates the expression of two genes, with normalized expression levels g_1 and g_2; in addition, the product proteins of gene 1 regulate the expression of gene 2. Within this class of networks, we can calculate the information transmission following the outline above, and then search for optimal values of all the parameters.

We find that the general problem has many well defined local optima, corresponding to different topologies of the network, as shown in Fig. 3. With just two target genes and a single input, the constraint of considering feed forward networks means that the *locations* of the arrows in the graph describing the network are fixed, but it is conventional to distinguish different topologies based on the signs associated to the arrows, that is whether a transcription factor acts as an activator (A) or repressor (R), leaving us with multiple possibilities. For convenience we describe, for example, a network in which the input activates both target genes, while gene 1 represses gene 2, as $A_1A_2-R_{12}$.

Our intuition is that the role of interactions is to reduce redundancy, and this is borne out by the solutions associated to each topology. In particular, if the input activates both targets, then having one target activate the other can't help to reduce redundancy, and correspondingly we find that in the $A_1A_2-A_{12}$ topology the interaction between the target genes is actually driven to zero by the optimization, and this is true

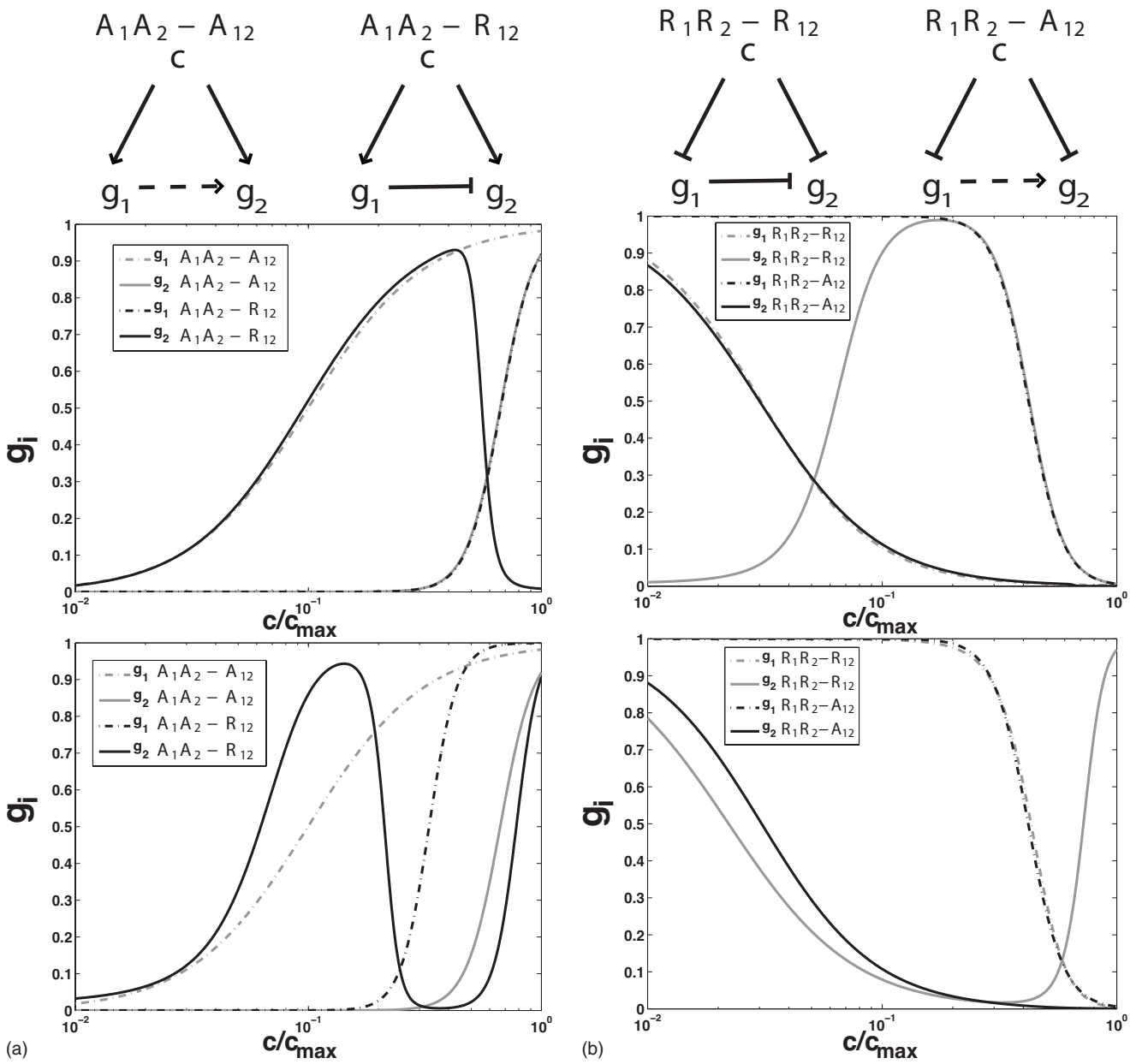

FIG. 3. Locally optimal networks for four topologies of the two gene feed forward networks, with $c_{max}/c_0=10$. Top panels are for the case of the Hill model, Eq. (19), and bottom panels are for the MWC model, Eq. (24). As explained in the text, the optimal parameters for the $A_1A_2-A_{12}$ and the $R_1R_2-A_{12}$ networks correspond to zero interaction between the target genes so these solutions also provide a reference for input/output relations in the optimal noninteracting networks. Note that the noninteracting solutions are very similar for Hill and MWC, while the optimal interacting networks are qualitatively different.

whether we use the Hill model of regulation or the MWC model (top and bottom left panels of Fig. 3). The same intuition suggests that the $R_1R_2-A_{12}$ topology should also be driven to zero interaction, and this too is what we find (right panels of Fig. 3). As an aside, these comparisons also verify that, in the noninteracting case, the Hill and MWC models achieve very similar input/output relations once we find the parameters that optimize information transmission.

The networks, which are optimized with nontrival interactions, are the ones where one target gene represses the other, $A_1A_2-R_{12}$ and $R_1R_2-R_{12}$. For the Hill model, both of these networks are optimized when one of the target genes is expressed only in a finite band or stripe of input concentrations, so that if we monitor only the expression level g_2, we see a nonmonotonic dependence on the input transcription factor. If we think of the input concentrations as varying along one axis of a two dimensional space—as they do in the genetic networks controlling embryonic development [17–20]—then this nonmonotonic behavior means that one of the genes will be expressed in a "stripe" through the middle of the spatial domain. The other gene, which is itself a repressor, is expressed only at the extremes of the concentration range, either at high or low input concentrations depending on whether the input is an activator or repressor, respectively. As in the case of noninteracting target genes, all of this structure depends on the available dynamic range of

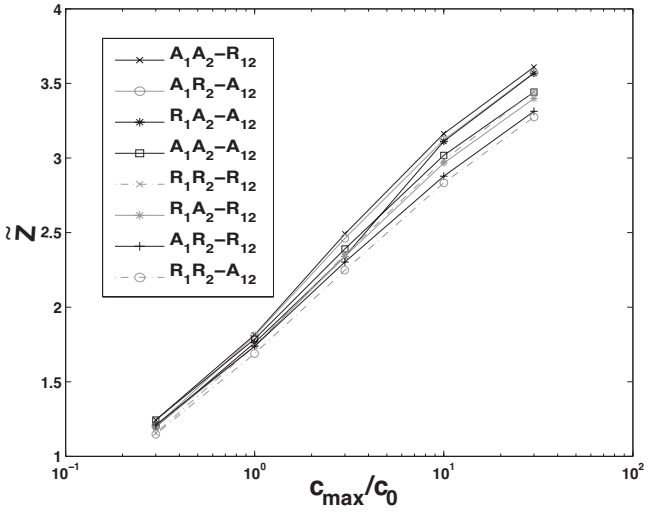

FIG. 4. Optimal information transmission in different network topologies, with Hill model regulation. The information capacity of a network depends on Z, from Eq. (38), which in turn is proportional to the number of independent copies of the output molecules, N_g. Here we plot $\tilde{Z}=2\pi eZ/N_g$ as a function of the maximal concentration of input transcription factor, $c_{\max}/c_0$.

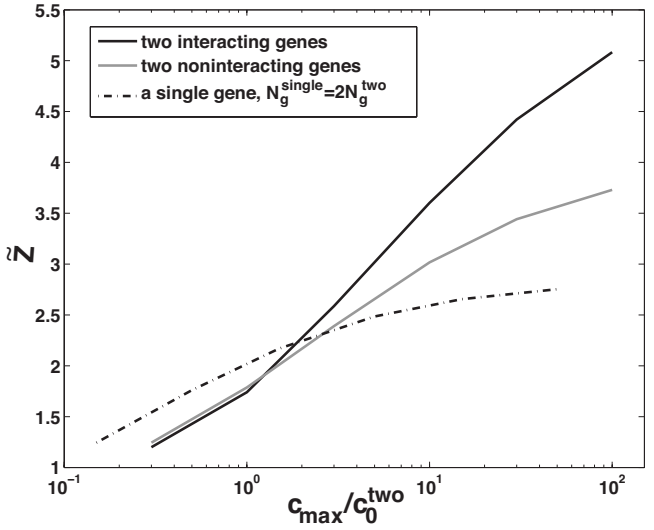

FIG. 5. One target or two? We show the dependence of $\tilde{Z}$ on the maximum concentration of input molecules, in the case of Hill regulation functions, comparing an interacting two gene network, a noninteracting two gene network, and a single output gene with twice the number of molecules. Note that to make the comparison meaningful, we have to be careful about the definition of c_0, since this concentration scale includes a factor of N_g; here we normalize to c_0 for the two gene networks.

inputs. As we make $c_{\max}/c_0$ smaller, the optimal strength of the interactions becomes smaller, and the networks that optimize information transmission are more nearly noninteracting; if $c_{\max}/c_0$ is sufficiently small, the optimal solution again is for all targets to be completely redundant, allowing the system to make multiple independent measurements of the same signal, rather than using the different targets to respond to different portions of the input dynamic range.

We know from the analysis of the noninteracting system that suboptimal settings of all the parameters really do incur large penalties, substantially reducing information transmission [16]. This remains true for the interacting case, in that if we choose a topology but set the parameters arbitrarily, there is a significant loss of information. On the other hand, once we choose parameters optimally, the different topologies have very similar information capacities, at least in the case of two target genes, as shown in Fig. 4. To set a scale for these results on information transmission, we could ask whether there really is an advantage to having two target genes. After all, noise is reduced by making more molecules, and with two targets the maximal number of output molecules will be increased. A useful comparison, then, is with a system that has just one target gene, but can generate twice as many output molecules, and this is shown in Fig. 5. We see that, as the structure of the optimal networks collapses at low values of $c_{\max}/c_0$, so too does the advantage of having two separate target genes. At very low concentrations of the input transcription factors, it really is better to have one target with twice the number of molecules, but as soon as there are enough molecules available to favor two target genes, the interactions between these genes become important in building the networks that optimize information flow.

One of the most interesting results of the optimization problem is that the Hill and MWC models, which are almost indistinguishable in the noninteracting case, are driven to qualitatively different solutions in the interacting case. The basic idea of how interactions relieve redundancy can be understood in a limit where each gene is either fully expressed ($\bar{g}=1$) or completely turned off ($\bar{g}=0$). For the noninteracting network, for example, with the A_1A_2 topology, the genes "turn on" in sequence, and so the system accesses states 00, 01, and 11 in order as the input concentration is increased. In the interacting network $A_1A_2-R_{12}$, the optimal Hill model (cf. Fig. 3, top left panel) accesses state 00, 01, and then switches to 10, not quite achieving the state 11. In contrast, the MWC model, once optimized, accesses all the states in sequence, $00\rightarrow01\rightarrow10\rightarrow11$, as the input concentration is increased, as we can see from the lower left panel of Fig. 3; the situation is reversed in the case of $R_1R_2-R_{12}$. We can understand how this happens, because in the MWC model each binding event shifts the equilibrium between the active and inactive states; thus there is a competition on gene 2 between the activating effect of the input and the repressing effect of gene 1, and this competition continues throughout the range of input concentration. In the MWC model, by adjusting parameters, it is possible that the final increase in input concentration up to $c_{\max}$ overwhelms the repressive effect of gene 1, and indeed this is what happens at the parameter values that optimize information transmission.

The difference between the Hill and MWC models indicates that the relatively microscopic physics of protein-DNA interactions which underlies transcriptional regulation can be "felt" at a more macroscopic, phenomenological level. In this spirit, we should also ask if it matters that we assume each protein can act only as an activator or a repressor, but not both. It is easy to imagine the alternative in which, by proper placement of the binding sites, the transcription factor can activate one gene by recruiting the polymerase, but can

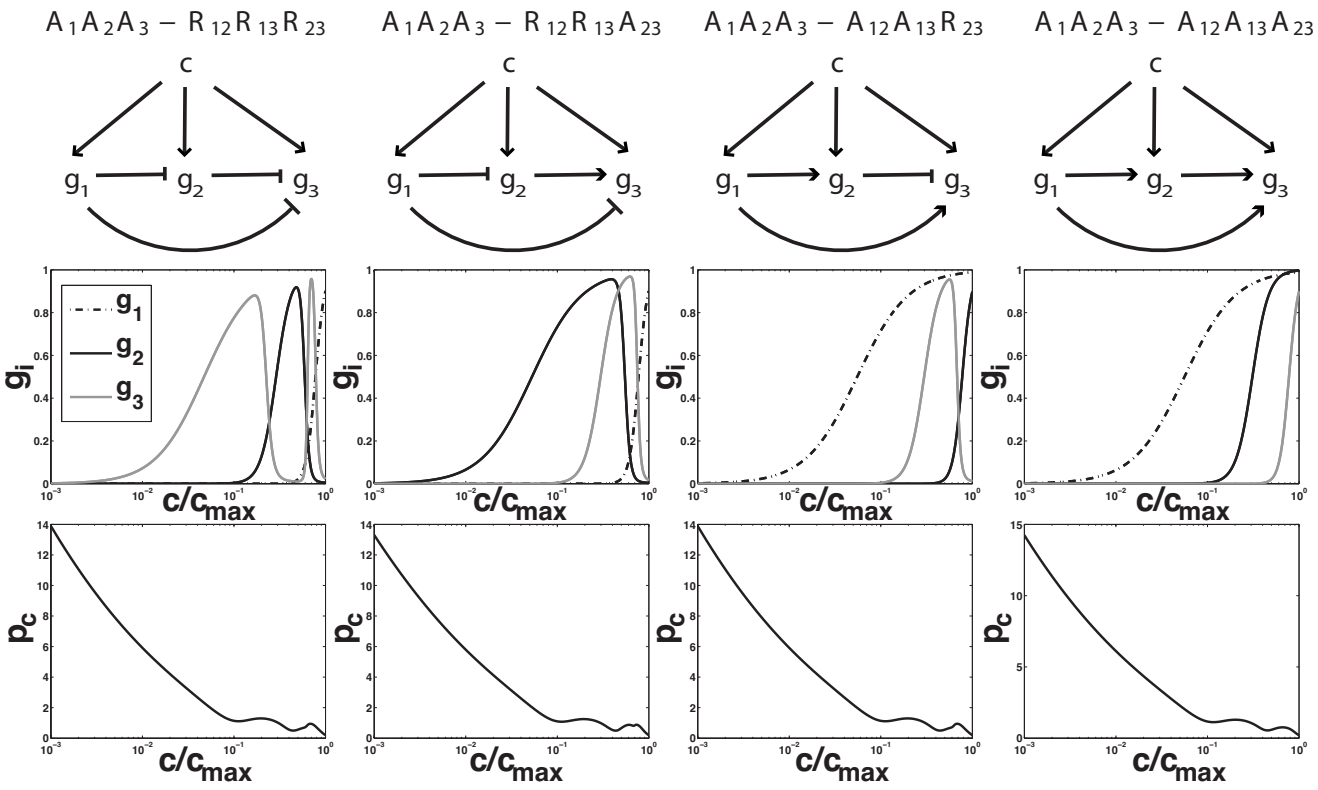

FIG. 6. Locally optimal networks with the three target genes, with the input acting as an activator. Regulation is according to the Hill model, Eq. (19), and we choose parameters to maximize information transmission with $c_{max}/c_0=10$. Top row shows the network topologies, and the middle row shows the average expression levels of the different target genes, as in Fig. 3. The bottom panel shows the distribution of input concentrations that matches the optimal networks, from Eq. (37).

repress other genes by occlusion. We have analyzed these mixed networks, and find that the optimal information transmission in these cases is intermediate between the "pure" networks considered thus far. Thus, considering these additional topologies does not change the global picture of the problem, except to open the possibility of yet more local optima.

B. Three target genes

The analysis of networks with three target genes confirms and amplifies the lessons learned in the two gene case. Because the space of possible networks is much larger, we once again restrict individual proteins to act only as activators or repressors, and give explicit results only for the case of the Hill function models of regulation, Eq. (19); results are shown in Figs. 6 and 7. As with two genes, having the target genes activate one another cannot reduce the redundancy that is generated in the noninteracting networks, and hence the $A_1A_2A_3-A_{12}A_{13}A_{23}$ topology networks are driven back to the noninteracting $A_1A_2A_3$ when we solve the optimization problem, and similarly for $R_1R_2R_3-A_{12}A_{13}A_{23}$ Topologies that include some repressive interactions generate nontrivial solutions, with progressively richer structure the more repression we allow.

The signature of the repressive interactions in the case of two targets was the emergence of nonmonotonic dependencies of the expression levels on the input transcription factor concentration. In the case of three targets, the parameters which maximize information transmission again lead to nonmonotonicity. In the case where the input is an activator (Fig. 6), allowing for one repressive interaction ($A_1A_2A_3-A_{12}A_{13}R_{23}$) leads to one gene having a nonmonotonic response. If we allow two repressive interactions ($A_1A_2A_3-R_{12}R_{13}A_{23}$) then two genes have nonmonotonic responses. Finally, with the maximum of three repressive interactions ($A_1A_2A_3-R_{12}R_{13}R_{23}$) one of the two nonmonotonic responses becomes yet more complex, with two "stripes" of expression in different ranges of the input concentration. In the case of an input repressor, we see a similar pattern (Fig. 7).

A striking feature of the networks with two target genes was that local optima with different topologies were nearly degenerate. This is less true for the case of three targets—the spread in capacities associated with the different topologies is larger, and the network $A_1A_2A_3-R_{12}R_{13}R_{23}$ is the clear global optimum. Corresponding to the larger spread in capacities, the enhancement of information transmission by interactions is larger, and in the difference in capacity between the Hill and MWC models also is larger in the three gene networks (Fig. 8). Including the possibility that individual transcription factors could act as both activators and repressors does not change these conclusions.

As a caveat, we note that we have assigned the same maximal concentration to all the transcription factors in the network. This is equivalent to assuming that all these molecules come at the same cost to the organism. One could

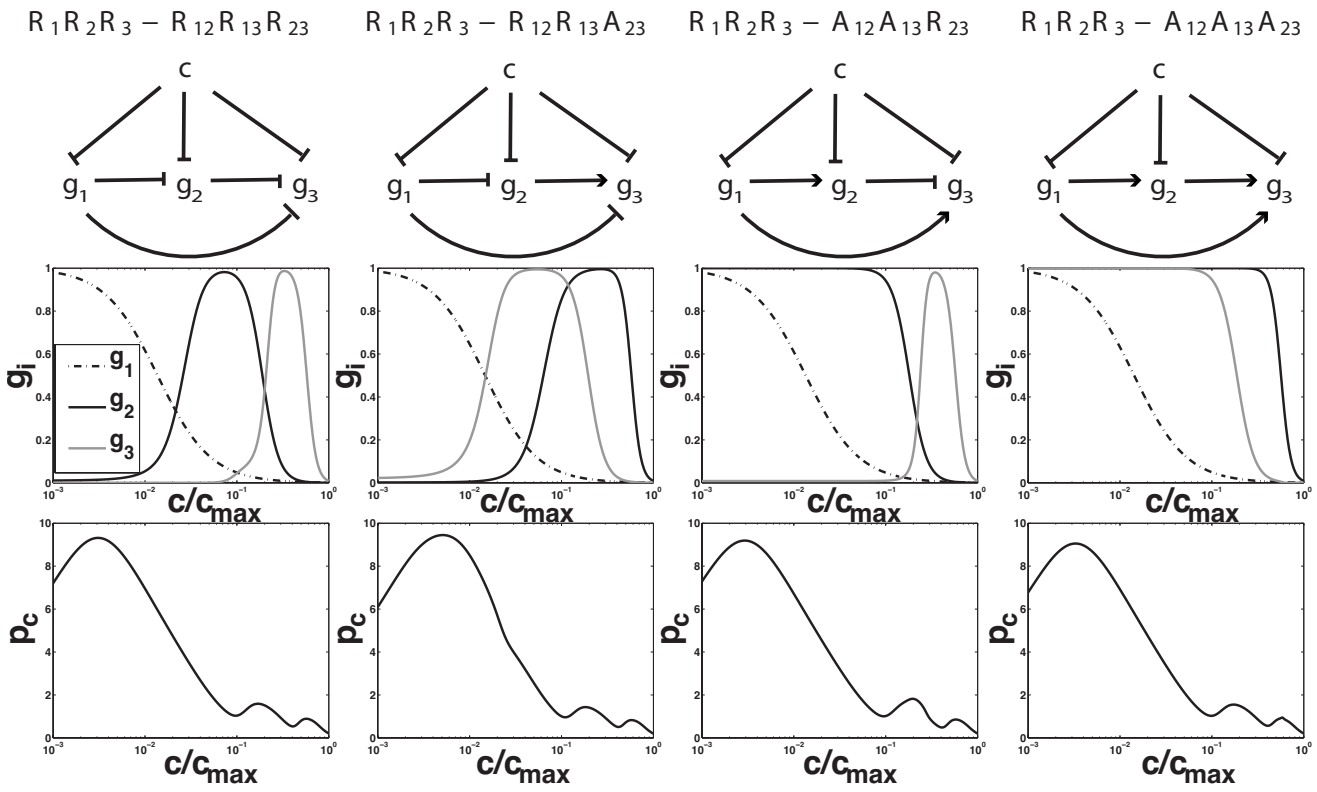

FIG. 7. Locally optimal networks with the three target genes, with the input acting as a repressor; all else as in Fig. 6.

imagine that there are significant differences between the molecules, and hence that one could be expressed at higher levels than the others for the same cost. If this symmetry is broken explicitly, then the ordering of the local optima associated to different network topologies can be shifted. While it is satisfying to identify a global optimum, it probably thus still makes sense to think of the problem as having many local optima that might all be relevant in different biological contexts. An interesting feature of these different optima is that, despite the changing structure of the input/output relations for the target genes, the distribution of input concentrations that optimizes information transmission [cf. Eq. (37)] is almost the same for all the cases where the input is an activator (bottom panels of Fig. 6) and again for all the cases where the input is a repressor (bottom panels of Fig. 7). Although maximizing information flow requires matching of the input distribution to the characteristics of the regulatory network [6,7], it seems that once the parameters of the network itself have been optimized, the solution to the matching problem has a more universal structure, with almost the same distribution providing the best match to several different local optima.

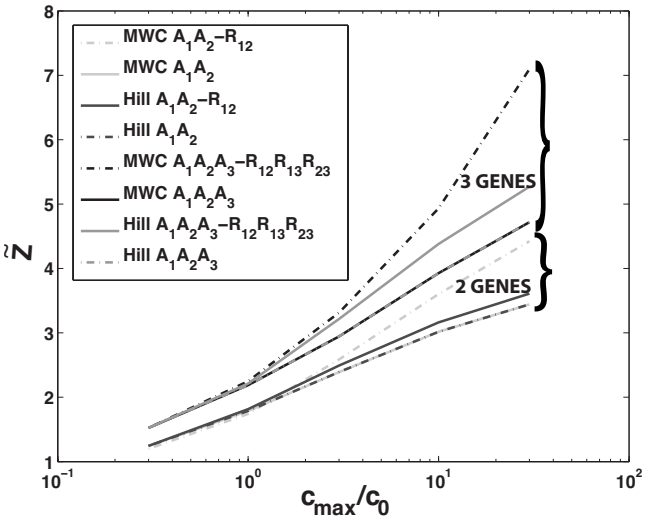

FIG. 8. Dependence of $\tilde{Z}$ on the maximum available protein concentrations $c_{\max}/c_0$ for two and three gene networks. We show results for the noninteracting cases, and for the globally optimal topologies, using both the Hill and MWC models of regulation.

Finally, we consider the parameter values in these optimal networks. For simplicity we focus on the globally optimal solutions with Hill model regulation functions for two $(A_1A_2-R_{12})$ and three $(A_1A_2-R_{12}R_{13}R_{23})$ target genes, plotting in Fig. 9 the evolution of the different constants K_i and K_{ij} as a function of the maximal input concentration $c_{\max}/c_0$. As noted above, the optimal interacting networks collapse onto the noninteracting case at small $c_{\max}$; for repressors, interactions become negligible if K_{ij} becomes large, and this is what we see at small $c_{\max}$. At the opposite extreme, as $c_{\max}/c_0$ becomes large, the optimal K_i become nearly constant fractions of $c_{\max}$, while the interactions become stronger and stronger even up to $c_{\max}/c_0 \sim 10^2$. Put another way, the concentrations at which the target genes are turned on by the input become distributed at fixed fractions of the available dynamic range, while the points at which the targets are turned off by their mutual repression evolves as a function of

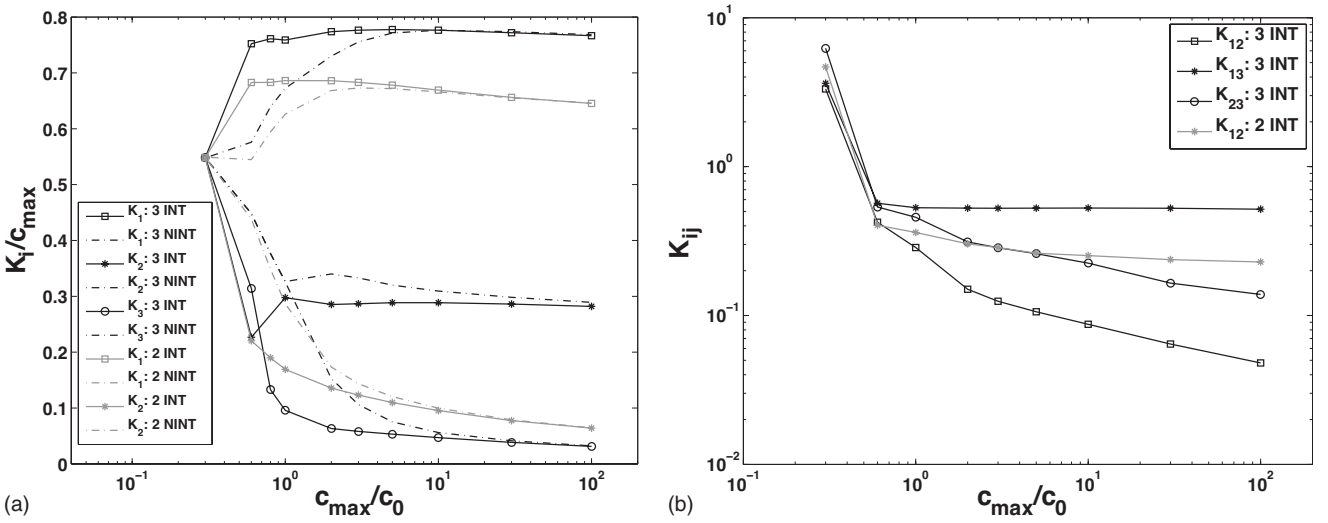

FIG. 9. Optimal parameters for Hill regulatory functions in the globally optimal topologies of two ($A_1A_2-R_{12}$ gray lines) and three ($A_1A_2A_3-R_{12}R_{13}R_{23}$ black lines) gene networks compared to the parameters of the optimal noninteracting networks (A_1A_2,$A_1A_2A_3$). Parameters are presented as a function of the maximum concentration of available input proteins c_{max}/c_0.

the maximal input concentration. Further, in the interacting networks, the difference between two and three genes is not simply an extra target, but a readjustment of the interactions to make optimal use of this extra output channel.

IV. DISCUSSION

The expression levels of target genes provide information about the concentration of the transcription factors which provide input to the regulatory network. Our goal here has been to understand how the structure of the network can be adjusted to optimize this information transmission. Earlier work [16] addressed this problem in the limit where the target genes do not interact, so the input is simply "broadcast" to multiple targets; here we considered the role of interactions within a network of target genes, limiting ourselves, for simplicity, to feed forward structures. We have found that there are an exponentially large number of locally optimal networks, each with parameters tuned to a well defined point that depends only on the maximal number of available molecules. These optimal networks have two components. As in the noninteracting case, the affinities of the input transcription factor for its multiple targets must be adjusted to balance the need to use the full dynamic range of inputs against the need to avoid noise at low input concentrations and output expression levels. Unique to the interacting case, mutual repression among the target genes is tuned to reduce the redundancy among the network's outputs.

The problem of redundancy reduction has a long history in the context of neural coding, dating back to the first decade after Shannon's original work [52–54]. As first emphasized for the retina, nearby neurons often receive correlated signals; since the dynamic range of neural outputs is limited, this redundancy compromises the ability of the system to transmit information. The solution to this problem is lateral inhibition— neighboring neurons that receive correlated inputs inhibit one another, so that each neuron transmits something which approximates the difference between its input and that of its neighbors. Lateral inhibition, sometimes mediated through a more complex network, is a common motif in neural circuitry, and in the retina it is expressed by the "center-surround" organization of the individual neurons' receptive fields. A more careful analysis shows that there is a tradeoff between redundancy reduction and noise reduction, so that the extent of inhibitory surrounds should be reduced, even to zero, as background light intensity is reduced and photon shot noise is increased, and this is in semiquantitative agreement with experiment [55,56]. The relation of these ideas in neural coding to our present discussion of genetic networks should be clear. Although there are many questions about how these ideas connect more quantitatively to experiment, it is attractive to think that similar physical principles, and even analogous implementations of these principles, might be relevant across such a wide range of biological organization.

There are relatively few transcriptional regulatory networks that have been characterized to the point of comparing quantitatively with the sorts of models we have explored here. In contrast, there is a growing literature on the qualitative, topological structure of these networks. In particular, a number of groups have focused on the local "motif" structure of large networks, searching for patterns of interaction that are over-represented relative to some randomized ensemble of networks [57,58]. One such motif is the feed forward loop, which essentially captures the whole set of feed forward networks with two target genes that we have considered here. In the language of Fig. 1(b), much of the discussion about the importance of this motif has centered on the nature of the transformation from the input concentration c to the expression level of gene 2. By choosing coherent (in our notation, $A_1A_2-A_{12}$) or incoherent ($A_1A_2-R_{12}$) loops, the system can achieve different temporal dynamics as well as nonmonotonic input/output relations [59–64]. Since we have considered system in steady state, it is this last point which is significant for our discussion. Our results comparing Hill and

MWC models of regulation emphasize, however, that it is not just the topology of the network which is important for the structure of these input/output relations.

One of the best opportunities for quantitative comparison with experiment is in the genetic networks controlling the early events of embryonic development. In these systems, the primary morphogen molecules are sometimes transcription factors, so that we can literally see the variations in input concentration laid out in space [18–20]. Along the anterior-posterior axis of the *Drosophila* embryo, for example, the maternally provided morphogen Bicoid varies in concentration almost exponentially, so that the long axis of the embryo is essentially a logarithmic concentration axis [65,66]. Bicoid is a transcription factor that provides input to the network of gap genes, *hunchback*, *krüppel*, *giant*, and *knirps*. Although Bicoid is an activator, these genes exhibit nonmonotonic expression levels, forming stripes along the length of the embryo, and these nonmonotonicities are mediated by mutual inhibitory interactions [67–71]. Qualitatively, the structure of the optimal networks that we derive here (especially the globally optimal $A_1A_2A_3-R_{12}R_{13}R_{23}$) resembles the gap gene network, and the expression profiles along the log c axis are reminiscent of the spatial profiles of gap expression. Generalizations of the experiments in Ref. [14] should make it possible to map the input/output relations more quantitatively, and perhaps even to measure the interactions directly by detecting the predicted correlations among the noise in the expression levels of different gap genes. To make truly quantitative comparisons, however, we need to solve our optimization problem allowing for networks with feedback.

ACKNOWLEDGMENTS

We thank J. Dubuis, T. Gregor, W. de Ronde, T. Mora, E. F. Wieschaus, and especially C. G. Callan for helpful discussions. Work at Princeton was supported in part by NSF under Grant No. PHY–0650617, and by NIH under Grant Nos. P50 GM071508 and R01 GM077599. G.T. was supported in part by NSF under Grant Nos. DMR04–25780 and IBN–0344678, and by the Vice Provost for Research at the University of Pennsylvania. W.B. also thanks his colleagues at the Center for Studies in Physics and Biology at Rockefeller University for their hospitality during a portion of this work.

APPENDIX

This appendix collects some technical details in the calculation of the inverse covariance matrix $\mathcal{K}$, following the framework established in Sec. II D. We start by considering the case of one input transcription factor at concentration c controlling the expression of two output genes, with normalized expression levels g_1 and g_2. In this case the Langevin equations of motion describing the network are

$$\tau\frac{dg_1}{dt}=f_1(c)-g_1+\xi_1(t), \tag{A1}$$

$$\tau\frac{dg_2}{dt}=f_2(c,g_1)-g_2+\xi_2(t). \tag{A2}$$

We linearize the equations of motion around the steady state solutions $\{\bar{g}_1,\bar{g}_2\}$ and Fourier transform to obtain Eq. (46), where the matrix $\hat{A}$ is explicitly given by

$$\hat{A}(\omega)=\begin{bmatrix}1-i\omega\tau & 0\\ -\phi_{21} & 1-i\omega\tau\end{bmatrix}, \tag{A3}$$

and hence

$$\hat{A}^{-1}(\omega)=\frac{1}{(1-i\omega\tau)^2}\begin{bmatrix}1-i\omega\tau & 0\\ \phi_{21} & 1-i\omega\tau\end{bmatrix}. \tag{A4}$$

The covariance matrix then can found, from Eq. (52)

$$\mathcal{K}^{-1}=\int\frac{d\omega}{2\pi}\{\hat{A}^{-1}(\omega)\hat{N}[\hat{A}^{-1}(\omega)]^{\dagger}\}=\int\frac{d\omega}{2\pi}\frac{1}{(1-i\omega\tau)^2}\begin{pmatrix}1-i\omega\tau & 0\\ \phi_{21} & 1-i\omega\tau\end{pmatrix}\begin{pmatrix}N_{11} & 0\\ 0 & N_{22}\end{pmatrix}\frac{1}{(1+i\omega\tau)^2}\begin{pmatrix}1+i\omega\tau & \phi_{21}\\ 0 & 1+i\omega\tau\end{pmatrix}, \tag{A5}$$

$$=\int\frac{d\omega}{2\pi}\frac{1}{[1+(\omega\tau)^2]^2}\begin{bmatrix}N_{11}[1+(\omega\tau)^2] & \phi_{21}N_{11}(1-i\omega\tau)\\ \phi_{21}N_{11}(1+i\omega\tau) & N_{22}[1+(\omega\tau)^2]+\phi_{21}N_{11}(1+i\omega\tau)\end{bmatrix}, \tag{A6}$$

$$=\int\frac{d\omega}{2\pi}\frac{1}{[1+(\omega\tau)^2]^2}\begin{bmatrix}N_{11}[1+(\omega\tau)^2] & \phi_{21}N_{11}\\ \phi_{21}N_{11} & N_{22}[1+(\omega\tau)^2]+\phi_{21}N_{11}\end{bmatrix}. \tag{A7}$$

The elements of N_{ij} are given explicitly by

$$N_{11}=\frac{\tau}{N_g}\left[\bar{g}_1(c)+\left(\frac{\partial\bar{g}_1(c)}{\partial c}\right)^2 c\right], \tag{A8}$$

$$N_{22}=\frac{\tau}{N_g}\left[\bar{g}_2(c)+\left(\frac{\partial\bar{g}_1(c)}{\partial c}\right)^2 c+\frac{\bar{g}_1}{c_{\max}}\phi_{21}^2\right], \tag{A9}$$

where we have chosen the maximum concentrations for all types of proteins to be equal to $c_{\max}$, and we have set $c_0=1$,

as described in the text. To finish the calculation, we need the integrals

$$\int \frac{d\omega}{2\pi}\frac{1}{[1+(\omega\tau)^2]} = \frac{1}{2\tau}, \tag{A10}$$

$$\int \frac{d\omega}{2\pi}\frac{1}{[1+(\omega\tau)^2]^2} = \frac{1}{4\tau}. \tag{A11}$$

Then

$$\mathcal{K}^{-1} = \frac{1}{2\tau}\begin{bmatrix} N_{11} & \phi_{21}N_{11}/2 \\ \phi_{21}N_{11}/2 & N_{22}+\phi_{21}N_{11}/2 \end{bmatrix}, \tag{A12}$$

$$\mathcal{K} = \frac{2\tau}{N_{11}(N_{22}+\phi_{21}N_{11}/2) - N_{11}^2\phi_{21}^2/4} \times \begin{bmatrix} N_{22}+\phi_{21}N_{11}/2 & -\phi_{21}N_{11}/2 \\ -\phi_{21}N_{11}/2 & N_{11} \end{bmatrix} \tag{A13}$$

Lastly, we substitute the obtained covariance matrix into the expression for the normalization constant Z, as obtained in Eq. (38)

$$Z = \int_0^{c_{\max}} dc \left[\frac{1}{2\pi e} \sum_{i,j=1}^{2} \frac{d\bar{g}_i(c)}{dc} \mathcal{K}_{ij}(c) \frac{d\bar{g}_j(c)}{dc} \right]^{1/2}. \tag{A14}$$

Since the information we are trying to maximize is, from Eq. (39), $I^* = \log_2 Z$, we now can optimize Z over the parameters of the regulation function. It is now clear, having set the concentration variables in terms of the natural scale $c_0 = 1$, that the only parameter in the problem is $c_{\max}$; all other parameters will be determined by maximizing Z.

In the case of a three gene network, the calculation is analogous to the two gene case. The matrix $\hat{A}$ is explicitly given by

$$\hat{A}(\omega) = \begin{bmatrix} 1-i\omega\tau & 0 & 0 \\ -\phi_{21} & 1-i\omega\tau & 0 \\ -\phi_{31} & -\phi_{21} & 1-i\omega\tau \end{bmatrix}, \tag{A15}$$

and the elements of the noise spectrum matrix are

$$N_{11} = \frac{1}{N_g}\left[\left(\frac{\partial \bar{g}_1(c)}{\partial c} \right)^2 c + \bar{g}_1 \right], \tag{A16}$$

$$N_{22} = \frac{1}{N_g}\left[\left(\left. \frac{\partial f_2(c,g_1)}{\partial c} \right|_{g_1=\bar{g}_1(c)} \right)^2 c + \phi_{21}^2 \frac{\bar{g}_1}{c_{\max}} + \bar{g}_2 \right], \tag{A17}$$

$$N_{33} = \frac{1}{N_g}\left[\left(\left. \frac{\partial f_3(c,g_1,g_2)}{\partial c} \right|_{\{g_i=\bar{g}_i(c)\}} \right)^2 c + \phi_{31}^2 \frac{\bar{g}_1}{c_{\max}} + \phi_{32}^2 \frac{\bar{g}_2}{c_{\max}} + \bar{g}_3 \right]. \tag{A18}$$

Following Eq. (52) to find the inverse covariance integral, analogously to the two gene case, requires evaluating the frequency integrals. The general form of these integrals is:

$$I_{ks} = \int \frac{d\omega}{2\pi}\frac{1}{(1+i\omega\tau)^k(1-i\omega\tau)^s} = \frac{1}{\tau}\int \frac{d\zeta}{2\pi}\frac{1}{(1+i\zeta)^k(1-i\zeta)^s} = \frac{1}{\tau}\binom{s+k-2}{k-1}2^{1-s-k}, \tag{A19}$$

When the dust settles, the elements of the inverse of the covariance matrix, $\mathcal{K}^{-1}$ from Eq. (52), are

$$\mathcal{K}_{11}^{-1} = N_{11}, \tag{A20}$$

$$\mathcal{K}_{12}^{-1} = \frac{1}{2}N_{11}\phi_{21}, \tag{A21}$$

$$\mathcal{K}_{13}^{-1} = \frac{1}{4}N_{11}[\phi_{21}\phi_{32}+2\phi_{31}], \tag{A22}$$

$$\mathcal{K}_{21}^{-1} = \frac{1}{2}N_{11}\phi_{21}, \tag{A23}$$

$$\mathcal{K}_{22}^{-1} = N_{22} + \frac{1}{2}N_{11}(\phi_{21})^2, \tag{A24}$$

$$\mathcal{K}_{23}^{-1} = \frac{1}{2}N_{11}[(\phi_{21})^2\phi_{32}+\phi_{31}\phi_{21}] + \frac{1}{2}N_{22}\phi_{32}, \tag{A25}$$

$$\mathcal{K}_{31}^{-1} = \frac{1}{4}N_{11}[\phi_{21}\phi_{32}+2\phi_{31}], \tag{A26}$$

$$\mathcal{K}_{32}^{-1} = \frac{1}{2}N_{11}[(\phi_{21})^2\phi_{32}+\phi_{21}\phi_{31}] + \frac{1}{2}N_{22}\phi_{32}, \tag{A27}$$

$$\mathcal{K}_{33}^{-1} = N_{33} + \frac{1}{2}[N_{22}(\phi_{32})^2 + N_{11}(\phi_{31})^2] + N_{11}\phi_{21}\phi_{32}\phi_{31} + \frac{3}{8}N_{11}(\phi_{21}\phi_{32})^2. \tag{A28}$$

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (Garland Science, New York, 2002).

[2] M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Press, New York, 2002).

[3] R. Wagner, *Transcription Regulation in Prokaryotes* (Oxford University Press, Oxford, 2000).

[4] W. Bialek and S. Setayeshgar, Proc. Natl. Acad. Sci. U.S.A. **102**, 10040 (2005).

[5] W. Bialek and S. Setayeshgar, Phys. Rev. Lett. **100**, 258101 (2008).

[6] G. Tkačik, C. G. Callan, Jr., and W. Bialek, Proc. Natl. Acad. Sci. U.S.A. **105**, 12265 (2008).

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

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

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

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

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

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

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

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

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

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

[17] L. Wolpert, J. Theor. Biol. **25**, 1 (1969).

[18] P. A. Lawrence, *The Making of a Fly: The Genetics of Animal Design* (Blackwell, Oxford, 1992).

[19] J. B. Gurdon and P. Y. Bourillot, Nature (London) **413**, 797 (2001).

[20] C. Nüsslein-Volhard, *Coming to Life: How Genes Drive Development* (Kales, Carlsbad, 2008).

[21] A comparable problem exists in the case of neurons. When multiple neurons respond independently to sensory inputs, the population of cells necessarily carries less information than would be expected by summing the information carried by the individual cells. This redundancy among conditionally independent events is quite general, and can be seen also for successive action potentials generated by the same neuron over time. See, for example, N. Brenner, S. P. Strong, R. Koberle, W. Bialek, and R. R. de Ruyter van Steveninck, Neural Comput. **12**, 1531 (2000).

[22] There are examples of transcription factors that act on one target as an activator and on another target as a repressor, but it is unusual that this happens with the factor acting alone. In eukaroytes, where transcription factors often form complexes, an activator without its partner can function as a repressor (Ref. [23]).

[23] A. H. Brivanlou and J. E. Darnell, Jr., Science **295**, 813 (2002).

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

[25] A. Mugler, E. Ziv, I. Nemenman, and C. H. Wiggins, IET Syst. Biol. **3**, 379 (2009).

[26] R. C. Yu, C. G. Pesce, A. Colman-Lerner, L. Lok, D. Pincus, E. Serra, M. Holl, K. Benjamin, A. Gordon, and R. Brent, Nature (London) **456**, 755 (2008).

[27] F. Tostevin and P. R. ten Wolde, Phys. Rev. Lett. **102**, 218101 (2009).

[28] P. Mehta, S. Goyal, T. Long, B. L. Bassler, and N. S. Wingreen, Mol. Syst. Biol. **5**, 325 (2009)..

[29] In bacteria, it often is assumed that the dominant term in the decay of protein concentrations is growth and division of the cells, which would make all lifetimes equal. More generally, we consider all transcription factors to operate under similar constraints, so that all of the network structure comes out of the optimization problem, rather than being imposed by hand.

[30] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 2007).

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

[32] J. Hasty, F. Isaacs, M. Dolnik, and D. McMillen, Chaos **11**, 207 (2001).

[33] J. Hasty, D. McMillen, F. Isaacs, and J. Collins, Nat. Rev. Genet. **2**, 268 (2001).

[34] J. Paulsson, Nature (London) **427**, 415 (2004).

[35] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Phys. Rev. E **79**, 031911 (2009).

[36] P. Mehta, R. Mukhopadhyay, and N. S. Wingreen, Phys. Biol. **5**, 026005 (2008).

[37] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, PLoS Biol. **4**, e309 (2006).

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

[39] P. J. Choi, L. Cai, K. Frieda, and X. S. Xie, Science **322**, 442 (2008).

[40] H. C. Berg and E. M. Purcell, Biophys. J. **20**, 193 (1977).

[41] G. Tkačik and W. Bialek, Phys Rev E **79**, 051901 (2009).

[42] We note that, in eukaryotes, the relevant volume for sensing the concentration of transcription factor proteins is the nucleus, however proteins are produced in the cytoplasm and then imported into the nucleus. Allowing for this partitioning, Ω becomes an effective volume, not the literal volume of the cell or nucleus.

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

[44] 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).

[45] T. Kuhlman, Z. Zhang, M. H. Saier, Jr., and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **104**, 6043 (2007).

[46] J. B. Kinney, G. Tkačik, and C. G. Callan, Jr., Proc. Natl. Acad. Sci. U.S.A. **104**, 501 (2007).

[47] A. V. Hill, J. Physiol. **40**, iv (1910).

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

[49] We have also studied the general model, with both K_{on} and K_{off} parameters. The optimum information transmission always occurs at $L \gg 1$, and the details of the optimal solution reproduce the behavior in the simpler limit considered here.

[50] 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).
[51] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[52] F. Attneave, Psychol. Rev. **61**, 183 (1954).
[53] H. B. Barlow, in *Sensory Mechanisms, the Reduction of Redundancy, and Intelligence*, Proceedings of the Symposium on the Mechanization of Thought Processes, edited by D. V. Blake and A. M. Utlley (HM Stationery Office, London, 1959), Vol. 2, pp. 537–574.
[54] H. B. Barlow, in *Sensory Communication*, edited by W. Rosenblith (MIT Press, Cambridge, 1961), pp. 217–234.
[55] J. J. Atick and A. N. Redlich, Neural Comput. **2**, 308 (1990).
[56] J. H. van Hateren, Nature (London) **360**, 68 (1992).
[57] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).
[58] U. Alon, Nat. Rev. Genet. **8**, 450 (2007).
[59] S. Mangan, A. Zaslaver, and U. Alon, J. Mol. Biol. **334**, 197 (2003).
[60] S. Mangan and U. Alon, Proc. Natl. Acad. Sci. U.S.A. **100**, 11980 (2003).
[61] S. Ishihara, K. Fujimoto, and T. Shibata, Genes Cells **10**, 1025 (2005).
[62] M. E. Wall, M. J. Dunlop, and W. S. Hlavacek, J. Mol. Biol. **349**, 501 (2005).
[63] S. Kalir, S. Mangan, and U. Alon, Mol. Syst. Biol. **1**, 2005.0006 (2005).
[64] S. Kaplan, A. Bren, E. Dekel, and U. Alon, Mol. Syst. Biol. **4**, 203 (2008).
[65] B. Houchmandzadeh, E. Wieschaus, and S. Leibler, Nature (London) **415**, 798 (2002).
[66] T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank, Cell **130**, 141 (2007).
[67] H. Jäckle, D. Tautz, R. Schuh, E. Seifert, and R. Lehmann, Nature (London) **324**, 668 (1986).
[68] R. Rivera-Pomar and H. Jäckle, Trends Genet. **12**, 478 (1996).
[69] L. Sanchez and D. Thieffry, J. Theor. Biol. **211**, 115 (2001).
[70] M. Levine and E. H. Davidson, Proc. Natl. Acad. Sci. U.S.A. **102**, 4936 (2005).
[71] D. Yu and S. Small, Curr. Biol. **18**, 868 (2008).