

Cross Talk and Interference Enhance Information Capacity of a Signaling Pathway

Sahand Hormoz*

Kavli Institute for Theoretical Physics, University of California–Santa Barbara, Santa Barbara, California

ABSTRACT A recurring motif in gene regulatory networks is transcription factors (TFs) that regulate each other and then bind to overlapping sites on DNA, where they interact and synergistically control transcription of a target gene. Here, we suggest that this motif maximizes information flow in a noisy network. Gene expression is an inherently noisy process due to thermal fluctuations and the small number of molecules involved. A consequence of multiple TFs interacting at overlapping binding sites is that their binding noise becomes correlated. Using concepts from information theory, we show that in general a signaling pathway transmits more information if 1), noise of one input is correlated with that of the other; and 2), input signals are not chosen independently. In the case of TFs, the latter criterion hints at upstream cross-regulation. We demonstrate these ideas for competing TFs and feed-forward gene-regulatory modules, and discuss generalizations to other signaling pathways. Our results challenge the conventional approach of treating biological noise as uncorrelated fluctuations, and present a systematic method for understanding TF cross-regulation networks either from direct measurements of binding noise or from bioinformatic analysis of overlapping binding sites.

INTRODUCTION

Accurate transmission of information is of paramount importance in biology. For example, in the process of embryonic development, crude morphogen gradients need to be translated into precise expression levels in every cell and sharp boundaries between adjacent ones (1,2). The embryo accomplishes this using a complex network of signaling molecules that regulate not only the expression level of the desired output gene but also each other. One simple strategy for increasing accuracy is to use multiple input signals. Indeed, frequently, the expression level of a single gene is controlled by multiple transcription factors (TFs) (take, for example, *bicoid* and *hunchback*, or *dorsal* and *twist*, in the *Drosophila* embryo (1,3,4)). These TFs, however, often have overlapping binding sites that result in interactions at binding and synergetic control of transcription (3,4).

Here, we suggest that interaction at the level of binding (interference) is related to the upstream network of TFs regulating each other (cross talk). Our main assumption is that the regulatory network is designed to optimize information transfer from the input (TF concentrations) to the output (gene expression level). This is a reasonable assumption in the case of development, where accurate positional information needs to be extracted from noisy morphogen concentrations (2).

First, we define the concept of a *cis*-regulatory network as a noisy communication channel, where the input encodes information by taking on a range of values, i.e., a morphogen gradient that carries positional information. Decoding this

information is subject to biological noise; for example, at the molecular level, the stochastic binding of morphogens to receptors makes an exact readout of their concentration impossible.

We show that in general, two input signals with correlated noise can transmit more information if they are not independent information carriers but are chosen from an entangled joint distribution, i.e., the concentration of one morphogen in a given cell is related to the concentration of another. Physically, this implies that the two inputs regulate each other upstream through cross talk. We demonstrate this by analyzing a simple model of two TFs competing for the same binding site. The competition at the binding site results in correlated binding/unbinding fluctuations. Solving for the optimal joint distribution of the input TF concentrations indicates that upstream, one is positively regulated by the other. Despite the increase in noise for each individual input from the competition, two interacting TFs can transmit more information than two noninteracting TFs because of 1), correlated noise in the inputs, and 2), an entangled optimal input distribution.

We suggest that this mechanism is consistent with the recurring strategy of the feed-forward motif, where one TF positively regulates another and both bind to partially overlapping sites that induce interactions. We confirm this claim by simulating the stochastic dynamics of all structural types of the feed-forward loop that are subject to correlated input noise. Three specific biological examples are discussed: joint regulation of the gene *Race* in the *Drosophila* embryo by the intracellular protein Smads and its target, *zen*; regulation of *even-skipped stripe 2* by *bicoid* and *hunchback*; and regulation of *snail* by *dorsal* and *twist*. Generalization to other forms of cross talk, such as cross-phosphorylation, and to other forms of

Submitted July 31, 2012, and accepted for publication January 22, 2013.

*Correspondence: hormoz@kitp.ucsb.edu

Editor: Edda Klipp.

© 2013 by the Biophysical Society
0006-3495/13/03/1170/11 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2013.01.033>



interference, such as the use of scaffold proteins, is also discussed.

The structure of this article is as follows. In the **Results** section, we first establish that in general, two input signals with correlated noise can transmit more information if their input joint distribution is not separable. We then consider a model of two TFs competing for the same binding site. We calculate analytically the noise correlations and the associated optimal input distribution. In some regimes, competing TFs outperform independent ones. Motivated by this finding, in the last subsection of the **Results**, we ask whether realistic gene regulatory modules can generate close-to-optimal input distributions and combine correlated inputs to maximize information transmission. We compute numerically the channel capacity for feed-forward loops, where joint regulation of the target gene is subject to generic correlated noise; noise correlations in these simulations are a parameter used to generalize the results beyond competing TFs. Relevant biological examples are considered in the **Discussion**.

Gene regulation as a communication channel

Regulatory networks in a cell are information-processing modules that take in an input, such as concentration of a nutrient, and generate an output in the form of a gene expression level. Information in the input is typically encoded as the steady-state concentration of a TF, c , which binds to the promoter site of the desired response gene and enhances or inhibits its transcription. At a molecular level, the process of binding is inherently noisy, subject to thermal agitations and low-copy-number fluctuations (5–7). The noise is captured through a probabilistic relationship between the TF concentration, c , and the gene expression level, g , $P(g|c)$. The detailed form of $P(g|c)$ depends on physical parameters such as binding and unbinding rates. We can think of this process as communication across a noisy channel (8). To alleviate the impact of the noise, various strategies can be adopted, such as limiting the input to sufficiently spaced discrete concentration levels, c_i , that result in nonoverlapping outputs. In many gene regulatory networks, spatial and temporal averaging of input signals are also used to reduce noise (9).

Shannon's channel coding theorem (10,11) tells us the maximum rate at which information can be communicated across a noisy channel, also referred to as the channel capacity. Throughout this work, we assume that gene regulatory networks are selected to optimize the rate of information transmission. This is a strong but reasonable assumption; for example, the cell will clearly benefit from a more accurate knowledge of the amount of nutrient in its environment. However, the cost of an optimal network can exceed the benefit of more accurate information. Here, we do not account for the cost of a network—the only metric for comparison is the channel capacity.

With knowledge of the nature of the noise in a channel, $P(g|c)$, it is possible to compute the probability distribution of the input signal, $P_{TF}^*(c)$, that maximizes the rate of information transmission. Essentially, this distribution tells the sender how often a particular TF concentration should be used for optimal transmission of information encoded in concentration. However, it does not tell the sender anything about the encoding and decoding schemes. This abstraction is useful, allowing us to compute the optimal input without having derived the optimal coding. However, the optimal coding might require input blocks of infinite size and complex codebooks, with little biological relevance.

Nevertheless, there are experimental observations consistent with the idea of regulatory systems maximizing information transmission rates. Tkacik et al. (12) have shown that experimental measurements of Hunchback concentration in early *Drosophila* embryo cells (9) show a distribution that closely matches the optimal frequency for the measured levels of noise in the system; with the system achieving 90% of its maximum transmission rate.

METHODS

Competing-TFs binding model

The fractional occupation of the binding site by TF i (n_i) satisfies the kinetic equation

$$\frac{dn_i(t)}{dt} = kc_i(1 - n_1 - n_2) - ln_i + \xi_{n_i} \quad i = 1, 2,$$

where k , c_i , and l are the on rate, TF concentration, and off rate, respectively. The Langevin noise term, ξ_{n_i} , introduces uncorrelated fluctuations: $\langle \xi_{n_i}(t) \rangle = 0$ and $\langle \xi_{n_i}(t)\xi_{n_j}(t') \rangle = D_i^2 \delta_{ij} \delta(t' - t)$. For independent TFs, the first term on the righthand side is modified to $kc_i(1 - n_i)$. For the results quoted and displayed (see Fig. 4) in the **Results** section 'Integration time and cooperativity', above stochastic differential equations were numerically integrated using the Euler-Maruyama method (13) discretized with $dt = 0.01$ for the parameters $k = 1$, $c_{1,2} = 0.01$, $l = 10^{-4}$, and $D_i^2 = kc_i(1 - \bar{n}_1 - \bar{n}_2) + \bar{l}n_i = 2\bar{l}n_i$. \bar{n}_i denotes the average steady-state value of n_i . The power spectral density is computed using the Wiener-Khinchin formula, $S_i(\omega) = 1/T \langle | \int_0^T n_i(t) e^{-i\omega t} dt |^2 \rangle$ (14).

Feed-forward-motif kinetic equations

The concentrations of Y and g are given by the kinetic equations (15–17)

$$\begin{aligned} \frac{dY}{dt} &= \beta_y f(X + \zeta_x, K_{xy}) - \alpha_y Y + \eta_y \\ \frac{dg}{dt} &= \beta_g F(X + \xi_x, K_{xg}; Y + \xi_y, K_{yg}) - \alpha_g g + \eta_g \end{aligned}$$

f can be an activator, $f(x, k) = (x/k)^H / (1 + (x/k)^H)$, or a repressor, $f(x, k) = 1 / (1 + (x/k)^H)$, where H is the Hill coefficient. K_{ij} is the regulation coefficient of gene j by TF i . For an AND-gate, $F(x, k_x; y, k_y) = f(x, k_x)f(y, k_y)$, and an OR-gate: $F = f_o(x; k_x, y, k_y) + f_o(y; k_y, x, k_x)$, where for an activator, $f_o(x; k_x, y, k_y) = (x/k_x)^H / (1 + (x/k_x)^H + (y/k_y)^H)$, and for a repressor, $f_o(x; k_x, y, k_y) = 1 / (1 + (x/k_x)^H + (y/k_y)^H)$.

The upper equation captures cross-regulation of TF Y by TF X . The output noise is captured by the Langevin term η_y . The input noise—from fluctuations in the readout of the TF X concentration due to binding fluctuations, diffusion noise, etc.—is captured by the phenomenological noise term ζ_x . The noise in cross-regulations is an extrinsic noise in the system, since our channel is defined as the joint regulation of gene g by TFs X and Y (lower equation).

The intrinsic noise contains the output noise in synthesis and degradation fluctuations—shot noise—in g , captured by the Langevin term η_g . The intrinsic input noise is due to fluctuations in the readouts of the TF concentrations X and Y , which determine the synthesis rate of g through the Hill function F . The inputs of F fluctuate from the true concentrations of X and Y by stochastic terms ξ_x and ξ_y , respectively.

Phenomenological noise

We neglect the contribution of extrinsic noise, $\eta_y = \zeta_x = 0$; keeping it does not change the results qualitatively. The intrinsic input noise in the TF concentration readout satisfies $\langle \xi_{xy}(t) \rangle = 0$ and a general phenomenological form for TF concentration variance set by a constant term and one proportional to the TF concentration, since regardless of microscopic details, this noise stems fundamentally from finite, discrete, and fluctuating molecule numbers: $\langle \xi_x(t) \xi_x(t') \rangle = (\epsilon + qX)\delta(t' - t)$, $\langle \xi_y(t) \xi_y(t') \rangle = (\epsilon + qY)\delta(t' - t)$. The input noise can be correlated: $\langle \xi_x(t) \xi_y(t') \rangle = \rho q \sqrt{XY} \delta(t' - t)$, where ρ is the noise correlation coefficient, which is assumed, for simplicity, to be independent of the TF concentrations. A more complex structure for ρ —for instance, with concentration dependence as in the case of competing TFs—does not change the results qualitatively. ϵ is a small constant that ensures a minimum noise of one TF molecule/cell. For the output noise, $\langle \eta_g(t) \rangle = 0$ and $\langle \eta_g(t) \eta_g(t') \rangle = (\epsilon + qg)\delta(t' - t)$.

Numerical simulations

The initial conditions are at time $t = 0$, $Y = 0$ and $g = 0$. $X = c_1$ for all times $t \geq 0$. The above stochastic differential equations were numerically integrated using the Euler-Maruyama method (13) from $t = 0$ to $t = 10$ discretized with $dt = 0.001$. Output statistics were gathered after the steady state was reached, the last 3000 time steps, for 1000 runs. The numerical integration was implemented in MATLAB R2011a.

Input-distribution optimization

Output distribution, $P(g|c_1)$, was computed for 30 values of c_1 equally spaced in the log scale from $\log(c_1) = -1$ to $\log(c_1) = 1.5$ (see [Model parameters](#) below). Discretization captures spatial averaging of the diffusing inputs (exponentially decaying from the source) by the cells (18). Using a higher resolution over the input range did not increase the capacity significantly, since optimizing over input distribution resulted in discrete (spaced out) inputs. Constrained nonlinear optimization with a sequential quadratic programming method (19) was used to numerically optimize over the input distribution, $P_{in}(c_1)$, and compute channel capacity. This optimization was implemented in MATLAB R2011a.

Model parameters

The range of input and noise parameters was selected to match that of experimental measurements of the morphogen Hunchback in the *Drosophila* embryo (9). The conclusions above were unaffected by changing the parameters, as long as the feed-forward loops (FFLs) had dynamics with nontrivial steady states. However, for the figures and

numbers quoted in the text, the parameters used were $\alpha_g = \beta_g = 100$, $\beta_y = 10$, $\alpha_y = 1$, $K_{xy} = K_{xg} = K_{yg} = 1$, $H = 1$, $q = 1$, $\epsilon = 0.001$.

RESULTS

Noise correlations enhance capacity

First, we quantify how correlations in noise of multiple inputs enhance the rate of information transmission, following closely the approach taken by Tkacik et al. (20). Consider two TFs with concentrations c_1 and c_2 that regulate the expression level of a gene (denoted as g). These values can vary, for example, as a function of space, as in the case of morphogens along an embryo. The frequency of observing a particular concentration occurrence, c_1 and c_2 , is given by $P_{TF}(c_1, c_2)$. The entropy—or uncertainty—of the inputs is maximized when this distribution is uniform, or all concentrations are equally likely, which implies that the maximum amount of information is gained when the TF concentrations are determined precisely. Of course, our aim is to maximize not the entropy in $c_{1,2}$ but rather the information conveyed to the expression level, g .

The noise in the expression levels results in a distribution of g for fixed TF concentrations, $P(g|c_1, c_2)$. Equivalently, we can fix the expression level g and consider the corresponding distribution of TFs, $P(c_1, c_2|g)$, assuming that there is unique set of inputs for every value of g . The two distributions are related by Bayes' rule. The amount of information communicated from $c_{1,2}$ to g is given by the mutual information between the distributions of $c_{1,2}$ and g (10),

$$I(g; c_1, c_2) = - \int dc_1 dc_2 P_{TF}(c_1, c_2) \log P_{TF}(c_1, c_2) + \int dg P_{exp}(g) \times \int dc_1 dc_2 P(c_1, c_2|g) \log P(c_1, c_2|g), \quad (1)$$

where the distribution of expression level g is given by $P_{exp}(g) = \int dc_1 dc_2 P(g|c_1, c_2) P_{TF}(c_1, c_2)$.

We assume that the noise in $c_{1,2}$ for a fixed expression level g is small and distributed as a Gaussian around the mean value, $\bar{\mathbf{c}}(g)$,

$$P(c_1, c_2|g) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(\mathbf{c} - \bar{\mathbf{c}}(g))^T \Sigma^{-1}(\mathbf{c} - \bar{\mathbf{c}}(g))\right], \quad (2)$$

where $\mathbf{c} = (c_1, c_2)$, and Σ is the covariance matrix over the conditional probability for fixed g , or the noise covariance matrix, $\Sigma_{ij}(g) = \langle (c_i - \bar{c}_i(g))(c_j - \bar{c}_j(g)) \rangle$.

The small-noise approximation says that it is meaningful to think of a mean one-to-one input-output response, which is what is commonly measured in experiments. We expand around the mean response to the next order. The approximation, although strong, has been verified for a variety of

regulatory systems (see, for example, Bicoid-Hunchback in Gregor et al. (9) and Little et al. (21) or, for other examples, Raser and O’Shea (22), Newman et al. (23), and Rosenfeld et al. (24)), and it enables us to analytically calculate the optimal distribution. We will relax these assumptions later with a numerical approach. The mutual information under this approximation is given by

$$I(g; c_1, c_2) = - \int dc_1 dc_2 P_{TF}(c_1, c_2) \log P_{TF}(c_1, c_2) + \frac{1}{2} \int dc_1 dc_2 P_{TF}(c_1, c_2) \log \left(\frac{|\Sigma^{-1}(\bar{g}(\mathbf{c}))|}{4\pi^2 e^2} \right), \quad (3)$$

where Σ^{-1} is evaluated at the mean value of expression level \bar{g} corresponding to a given \mathbf{c} .

To find the channel capacity, Eq. 2 is optimized for the input distribution, $P_{TF}(c_1, c_2)$. With the probability distribution’s normalization constraint introduced using a Lagrange multiplier, the optimal distribution must satisfy

$$\frac{\delta}{\delta P_{TF}(c_1, c_2)} \left[I(g; c_1, c_2) - \lambda \int dc_1 dc_2 P_{TF}(c_1, c_2) \right] = 0. \quad (4)$$

The optimal input distribution in the small-noise approximation (Eq. 4) is given by

$$P_{TF}^*(c_1, c_2) = \frac{1}{2\pi e Z} \frac{1}{\sqrt{|\Sigma|}}, \quad (5)$$

where Z is the normalization constant.

The maximum mutual information, or channel capacity, for transmitting information from TF concentrations to expression level is

$$I^* = \log_2 Z = \log_2 \left[\frac{1}{2\pi e} \int \int_{c_{min}}^1 dc_1 dc_2 \frac{1}{\sqrt{|\Sigma|}} \right]. \quad (6)$$

We have constrained the input concentration to lie in the normalized range, $c_{1,2} \in [c_{min}, c_{max} = 1]$. The minimum concentration is set by the molecular nature of the input: a minimum of one input molecule per cell is required.

We can repeat the same calculation for one TF while neglecting the other, effectively ignoring the covariance of the noise (off-diagonal components of Σ). With no covariance, the noise distribution is separable, $P(c_1, c_2|g) = P(c_1|g)P(c_2|g)$. The optimal input concentration for TF₁ will be $P_1^*(c_1) \sim 1/\sqrt{\Sigma_{11}}$, and its channel capacity will be $I_1^* \sim \log \int dc_1 1/\sqrt{\Sigma_{11}}$; expression for the other TF will be similar.

For the simple case where Σ is independent of \mathbf{c} , the channel capacity of the two TFs can be decomposed into its individual and joint contributions,

$$I^* = I_1^* + I_2^* - \frac{1}{2} \log(1 - \rho^2), \quad (7)$$

where $I_{1,2}^*$ is the channel capacity of the TFs individually, and $\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$ is the noise correlation coefficient for TF concentrations. Accounting for noise correlation enhances the rate of information transmission. In fact, in the limit of perfect correlation, $\rho \rightarrow \pm 1$, the capacity is infinite. This is expected, since under the small-noise approximation and perfectly correlated noise, some combination of inputs is always noise-free. Noise-free continuous variables can transmit infinite information. Fig. 1 is a pictorial representation of how noise correlations are beneficial. Essentially, the information is encoded in a combination of the two inputs (such as their difference), which is subject to less noise.

In general, the optimal input distribution (Eq. 6) is not separable to individual components, namely,

$$P_{TF}^*(c_1, c_2) \neq P_1^*(c_1)P_2^*(c_2), \quad (8)$$

where $P_{1,2}^*$ is the marginal distribution for $c_{1,2}$. In a sense, $P_{TF}^*(c_1, c_2)$ is an entangled distribution, where the concentration of one TF determines the probability of observing a certain concentration of the other. Biologically, this hints at upstream interactions between the TFs, the form of which should be predictable from the nature of the noise correlations.

The above abstract results are not surprising. The more important question is whether noise can be correlated, i.e., $P(c_1, c_2|g) \neq P(c_1|g)P(c_2|g)$, for the physical process of binding and unbinding of multiple TFs to a promoter region. We demonstrate this below using competing TF modules.

Competing TFs

TFs regulate gene expression levels by binding to *cis*-regulatory regions on the DNA. The design of these regions is

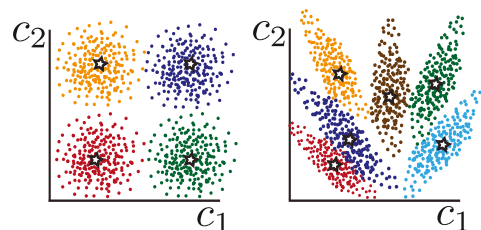


FIGURE 1 Benefit of correlated noise. (Left) Uncorrelated noise. Each color corresponds to a particular output response, g . Due to noise, many inputs (c_1, c_2) correspond to the same color output. For effective signaling, the outputs (and corresponding mean inputs, marked as stars) must be sufficiently spaced to avoid ambiguity. In this picture, four different outputs can be reliably communicated, corresponding to two values of c_1 and c_2 which can be selected independently. (Right) The two inputs have correlated noise, as reflected in the ellipsoidal scatter of input points (c_1, c_2) corresponding to the same output. Six distinct outputs can be reliably communicated due to smaller spread of noise in one direction. However, the nontrivial tiling means that the six allowed values of each input cannot be selected independently.

highly complex in both prokaryotes and eukaryotes, with overlapping TF binding sites occurring frequently (15,25,26).

We write a simplified model of the two extremes of overlapping binding sites (Fig. 2). The dominant source of noise is assumed to be the intrinsic noise from fluctuations in binding/unbinding of TFs to the promoter; we address the validity of neglecting the diffusion noise in the TF concentration in the Discussion section. Details of RNAP assembly and transcription are coarse-grained to a simple TF binding picture. Nonetheless, we will show that this simple model captures the essential role of noise correlations in a regulatory network.

Following the approach of (27), let $n_{1,2}$ be the fractional occupation of the binding site by competing TF_{1,2}. $n_1 + n_2 < 1$ is the fractional occupation of the site by either TF. A binding event can occur only if the site is unoccupied, $1 - n_1 - n_2$ of the time.

$$\frac{dn_i(t)}{dt} = kc_i(1 - n_1 - n_2) - ln_i \quad i = 1, 2 \quad (9)$$

The binding rate (on rate) is proportional to the concentration of the TF, and the off rates are given by constant l . At thermal equilibrium, these two rates are related through the principle of detailed balance, $kc_1/l = \exp(F_1/k_B T)$, where F_1 is the free-energy gain in binding for TF₁, with a similar expression for TF₂. We rescale time so that $k = 1$.

Eq. 10 is a dynamical picture of the fractional occupation of the binding site by each TF. At steady state, the mean fractional occupation is denoted by $\bar{n}_{1,2}$. We incorporate thermal fluctuations by introducing small fluctuations in $F_{1,2}$, capturing thermal kicks of energy that result in binding/unbinding events by effectively changing the binding energy.

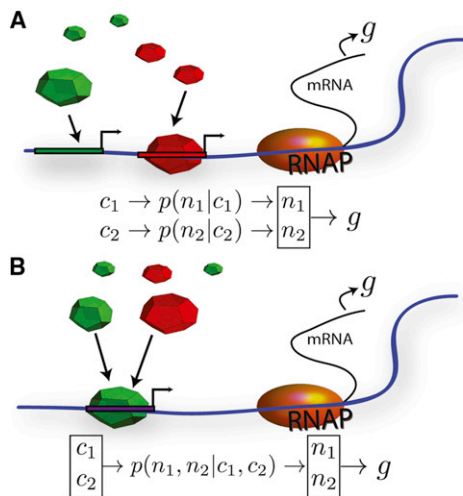


FIGURE 2 Independent versus competing TFs. (A) Nonoverlapping binding sites. Fractional binding-site occupations, $n_{1,2}$, are not correlated, nor is the noise in estimates of $c_{1,2}$. The expression level, g , is dependent on both inputs. (B) Overlapping binding sites. $n_{1,2}$ are dependent, resulting in correlated noise in estimating c_1 and c_2 .

We do not worry about fluctuations in c itself, i.e., from the extrinsic noise; the TF concentrations are the fixed inputs of the system and do not fluctuate. Fluctuations in the fractional occupation of the binding site, $n_{1,2}$ effectively introduce noise in the readout of the concentrations, $p(\mathbf{n}|\mathbf{c})$.

With this substitution and taking the Fourier transform, the linearized fluctuations around the mean $\delta n_{1,2}$ satisfy

$$\begin{pmatrix} \delta \tilde{F}_1 \\ \delta \tilde{F}_2 \end{pmatrix} = \frac{k_B T}{1 - \bar{n}_1 - \bar{n}_2} \begin{pmatrix} 1 + \frac{-i\omega + l_1}{k_1 c_1} & 1 \\ 1 & 1 + \frac{-i\omega + l_2}{k_2 c_2} \end{pmatrix} \begin{pmatrix} \delta \tilde{n}_1 \\ \delta \tilde{n}_2 \end{pmatrix} \quad (10)$$

where the tilde denotes the Fourier-transform, $\delta \tilde{n}_1(\omega) = \int_0^\infty \delta n_1(t) e^{i\omega t} dt$. In vectorial form, the relation becomes, $\delta \tilde{\mathbf{F}} = \Lambda \delta \tilde{\mathbf{n}}$.

Eq. 11 relates incremental fluctuations in the binding readout, $\delta \mathbf{n}$, with fluctuations in free energy, $\delta \mathbf{F}$. This is a linear response relation, with the free energy playing the role of the driving force (for details, see Bialek and Setayeshgar (27)). Using the fluctuation dissipation theorem (28), we calculate the power spectrum of noise in \mathbf{n} ,

$$S_n(\omega) = \frac{2k_B T}{\omega} \Im(\Lambda^{-1}), \quad (11)$$

with \Im denoting the imaginary part. From S , we can compute the covariance matrix,

$$\langle \delta \mathbf{n}^T \delta \mathbf{n} \rangle = \int_{-1/\tau_{int}}^{1/\tau_{int}} \frac{d\omega}{2\pi} S_n(\omega). \quad (12)$$

τ_{int} denotes the integration time of the site. For now, we assume $1/\tau_{int} \rightarrow \infty$ to compute the instantaneous fluctuations in the binding readout. Later, we will consider biologically relevant integration times. With proper normalization, we can compute the correlation coefficient (Fig. 3 A). The correlation coefficient is negative, since a more than expected occupation of the site by one TF will clearly result in less than expected occupation by the other.

Finally, we need to relate the noise in $\delta \mathbf{n}$ to the noise in the estimated TF concentrations. To do so, we account for the sensitivity of \mathbf{n} to the TF concentrations. For example, a very large c_1 results in $n_1 = 1$ with little noise. This readout, however, is not very sensitive to changes in c_1 and is not useful in detecting concentration changes. Let us define a matrix $\Omega_{ij} = \partial c_i / \partial \bar{n}_j$. The covariance matrix for the noise in TF concentrations is given by $\Sigma = \langle \delta \mathbf{c}^T \delta \mathbf{c} \rangle = \Omega \langle \delta \mathbf{n}^T \delta \mathbf{n} \rangle \Omega^T$.

In equating the covariance matrix in TF concentration to Σ (covariance matrix for a fixed g) of the previous section, we have introduced the extra assumption that the dominant

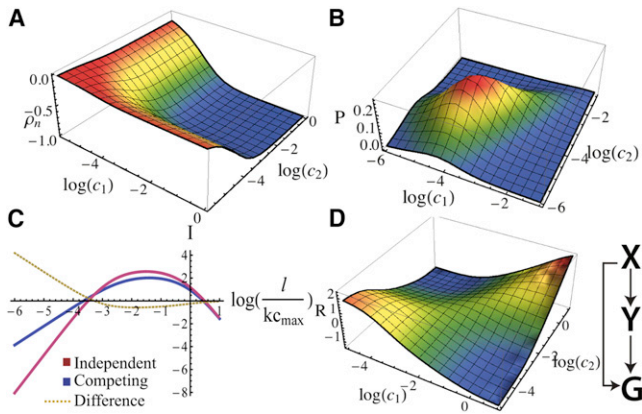


FIGURE 3 Competing TFs. (A) Correlation coefficient of readouts n_1 and n_2 , for $l = 10^{-4}$ and $c_{min} = 10^{-3}$ as a function of log input TF concentration. At high concentrations, a higher-than-expected readout of one TF implies a lower-than-expected readout of the other, resulting in a negative correlation coefficient. (B) The optimal input distribution for the same parameter values. (C) The channel capacity in bits for the interacting and noninteracting case of two TFs (blue and red curves, respectively) as a function of logarithm of rescaled l . The y-offset is arbitrary. The dashed curve denotes their difference. At biologically relevant $l = 10^{-4}$, interacting TFs have higher channel capacity. (D) The log likelihood of observing TF concentration (c_1, c_2) compared to what is expected from independent distributions. The likelihood of observing both TFs at either low or high concentrations together is significantly greater. This suggests that one TF positively regulates the other, feed-forward motif (right).

noise in the channel going from \mathbf{c} to \mathbf{g} is from the binding noise and not the expression level. Noise in \mathbf{g} is assumed negligible and need not be propagated backward and included in Σ . Since noise in \mathbf{g} is most commonly shot noise (29), this assumption is reasonable when expression levels are high. This also means that our results will not depend on the functional form of \mathbf{g} on \mathbf{c} (for the case when they do for one input, see (20,30,31)). We will relax this assumption below for the numerical simulations of the feed-forward loop.

We compute the optimal joint distribution of input concentration, $P_{TF}^*(c_1, c_2)$, by plugging the covariance matrix in Eq. 6 (Fig. 3 B). Moreover, Eq. 7 tells us the channel capacity, or the maximum information-transmission rate. Fig. 3 C plots channel capacity of two interacting TFs and two independent ones as a function of logarithm of off-rate $\log_{10}(l)$. The interacting TFs have a higher channel capacity in the biologically relevant regime where $l \sim 10^{-4}$ and $c_{min} \sim 10^{-3}$ (see below). This result does not depend sensitively on the lower bound of the TF concentration, c_{min} ; in fact, channel capacity is finite even when $c_{min} = 0$. The lower bound is enforced to ensure a minimum of one signaling molecule in the cell. The channel capacity has not increased simply because more signaling molecules are used in the interacting case. In fact, at these parameters, the mean input TF concentration, $\int \int dc_1 dc_2 (c_1 + c_2) P_{TF}^*(c_1, c_2)$, is $\sim 30\%$ less than that of the noninteracting channel.

The optimal joint distribution of input concentrations (Fig. 3 B) is entangled and no longer separable, $P_{TF}^*(c_1, c_2) \neq P_1^*(c_1)P_2^*(c_2)$. With an entangled distribution the system can explore degrees of freedom not present with two independent input distributions. In Fig. 3 D, we plot the log likelihood of observing joint concentration (c_1, c_2) compared to observing c_1 and c_2 independently from their marginal distributions. $R = \log_{10} P_{TF}^*(c_1, c_2) / P_1^*(c_1)P_2^*(c_2)$, where $P_1^*(c_1) = \int dc_2 P_{TF}^*(c_1, c_2)$ is the marginal distribution of TF₁, with a similar expression for TF₂.

Fig. 3 D implies that the two TFs are no longer passive and in fact cross-regulate each other. It is ~ 10 times less likely to observe one TF at a high concentration and the other at a low concentration simultaneously, compared to what is expected if they were independent. Similarly, it is ~ 10 times more likely to observe high concentrations of one TF if the concentration of the other is also high. This suggests that one TF positively regulates the other (Fig. 3 D, feed-forward motif).

Where does a biological system lie in the abstract parameter space sketched above? As noted, we have rescaled time so that $k = 1$ and measured concentration in units of $c_{max} = 1$. The only parameters left are the off-rate l and c_{min} . In a real cell, we expect a maximum of ~ 1000 TF molecules (or a dynamic range of 1–1000 TF molecules) in a volume of $\sim 1 \mu\text{m}^3$ (32). Hence, the minimum allowed concentration is $c_{min} = 10^{-3}$. A typical equilibrium constant of TF binding to DNA is $K_{eq} \sim 10^{10}\text{M}^{-1}$ (33). Putting all this together, we find $l \sim 10^{-4}$. It is possible, then, that a real biological regulatory system can transmit more information by incorporating overlapping binding sites and an upstream positive regulation between the TFs.

Integration time and cooperativity

To compute the channel capacity above, we used the instantaneous variance of the binding-site fractional occupation, $\langle \delta n(t)^2 \rangle$. In reality, however, a cell will integrate the occupation of the binding site for some time. The general theory proposed above is also valid in the limit $\tau_{int} \rightarrow \infty$. A longer integration time typically decreases both the variance and covariance by a factor $1/\tau_{int}$; the correlation coefficient ρ is unaffected. An entangled joint distribution of inputs is in general still more optimal than a separable one. However, the specific form of the frequency dependence of the noise can make the role of integration time more complicated. We examine information transmission in the above model for biologically relevant integration times.

The binding of a TF is a binary—on/off—signal for transcription of mRNA. The notion of a fractional occupation inherently assumes averaging over a series of binding and unbinding events. The first stage of integration is through transcription: the amount of time a TF is bound to DNA is approximately proportional to the amount of mRNA transcribed. Therefore, the lifetime of the mRNA, τ_e , sets the

transcriptional integration timescale. If mRNA lifetime is long, more mRNA molecules accumulate, resulting in a more precise value of the average time TF was bound. mRNAs in turn translate into protein; accumulation of proteins, with lifetime τ_g , is the second stage of integration.

Although generally $\tau_g \gg \tau_e$ (34), translation is a discontinuous process, with other sources of interruption besides binding fluctuations, i.e., chromatin remodeling and mRNA splicing in eukaryotes (35) and transcriptional bursting in prokaryotes (36). Naively, transcriptional integration removes fluctuations with frequencies higher than τ_e^{-1} ; translational integration has no frequency dependence—because it is punctuated—and simply reduces the variance of fluctuations by a factor of $1/\tau_g$. After the integration, the binding noise is estimated as

$$\langle \delta \mathbf{n}^T \delta \mathbf{n} \rangle \approx \frac{1}{\tau_g} \int_{-1/\tau_e}^{1/\tau_e} \frac{d\omega}{2\pi} S_n(\omega). \quad (13)$$

The power spectrum for interacting and noninteracting TFs is shown in Fig. 4 A from analytical calculations (Eq. 12) and numerical simulations of Eq. 10 (Methods). For a noninteracting TF, assuming $kc \gg l$, n fluctuates on the timescale $(kc)^{-1}$ (see derivation in Bialek and Setayeshgar (27)). Surprisingly, interacting TFs also show fluctuations at the longer timescale, l^{-1} ; refer to Fig. 4 B and the power spectrum in Fig. 4 C. However, the long wavelength fluctuations are almost perfectly anticorrelated between the two readouts

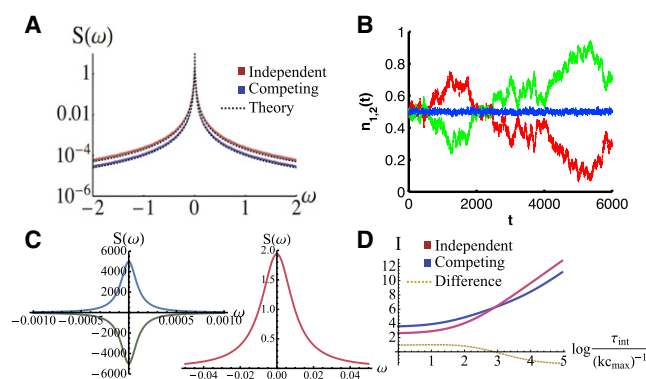


FIGURE 4 Integration time. (A) Power spectrum of the fluctuations in fractional binding, n , for both competing and independent TFs. The analytical calculation is in good agreement with simulation results (shown here for $k = 1$, $c = 0.01$, and $l = 10^{-4}$; see Methods). (B) A typical time-series of fluctuations in $n_{1,2}$ (red and green curves) exhibiting short-wavelength fluctuations (timescale $(kc)^{-1}$) and long-wavelength fluctuations (timescale l^{-1}). The long-term fluctuations are almost perfectly anticorrelated and can be removed by averaging the two inputs (blue curve). (C) Analytical power spectrum of the competing (left) and independent (right) TFs. Cross-power spectral density is also plotted for the competing case, which is negative because of the anticorrelations. The width of the power spectrum is on the order of kc for independent TFs but much narrower, on the order of l , for competing TFs. (D) Channel capacity as a function of the base-10 logarithm of normalized integration time for both types of TFs. Competing TFs outperform independent TFs for integration times up to $\sim 10^3 (kc_{max})^{-1}$.

n_1 and n_2 ($\rho \rightarrow -1$); when the readouts are combined, the remaining fluctuations have timescale $(kc)^{-1}$. Since the power spectrum of the competing TFs has a narrower width l , the integration time must be longer than l^{-1} for the noise to change substantially. TF dissociation rates can be slow, $l \sim 10^{-3} s^{-1}$ (33,37). A typical mRNA lifetime of minutes, $\tau_e \sim 10^2 s$ averages out the fast fluctuations at rate $k \sim 1 s$ (assuming a typical TF concentration of 100 nM) but does not filter the low-frequency fluctuations at rate l . The long protein lifetime, τ_g , typically many minutes to hours (34), averages over many binding and unbinding events.

We explicitly compute the channel capacity for competing and independent TFs as a function of the integration time (Fig. 4 D). The correlation coefficient of the fluctuations between the readouts does not diminish with increasing integration time. For low dissociation rates, $kc_{max}/l = 10^4$, competing TFs transmit more information than independent TFs up to an integration time of $10^3/kc_{max}$, $\sim 10^2 s$ using the above parameters, which is comparable to the biologically relevant integration time set by mRNA lifetime. We also explored the impact of cooperative binding of TFs by adding a Hill coefficient, λ , to the concentrations in Eq. 10 ($c_{1,2} \rightarrow c_{1,2}^\lambda$). The relative advantage of competing TFs disappears with increasing cooperativity. When $\lambda = 5$, the advantage of competing TFs disappears for any biologically relevant integration time. For $\lambda = 0.5$, however, the channel capacity is higher with competition—a larger increase compared with the uncooperative case—and persists to arbitrary large integration times, $\tau_{int} \rightarrow \infty$. It is conceivable that competing TFs may transmit more information than independent TFs in the limit of low dissociation rates or negative cooperativity for biologically relevant integration times; see the Discussion section for the importance of diffusion noise and its connection to cooperativity.

Feed-forward motif

The fact that interacting TFs have correlated noise is not surprising. The entangled optimal input distribution calculated above implied that one TF positively regulated the other upstream, reminiscent of a feed-forward motif. Naturally, the question arises whether a realistic biological model of feed-forward gene regulation can take advantage of correlated noise in the inputs. Can dynamical joint repression/activation of a target gene encode the signal in a combination of the two correlated inputs that is subject to less noise? Is it possible to optimize the input distribution using realistic gene regulatory modules? We answer these questions by numerically computing channel capacity of a FFL, where upstream, one TF regulates the other, and downstream, both jointly regulate the expression level of the target gene. Another purpose of the numerical approach is to relax the restrictive assumptions required for the above analytical derivations, in particular the small-noise approximation,

Gaussian form of the noise, one-to-one correspondence between input TF concentrations and the output expression level, and negligible output noise.

In the following analysis, we omit the details of how fluctuations in the readout of TF concentrations become correlated—one mechanism is overlapping binding sites (see above)—and simply introduce a general phenomenological model of input noise (Methods). Although many microscopic mechanisms can potentially generate correlated noise (interference), upstream cross-regulation of TFs is limited to certain well characterized gene-regulatory modules. Our purpose is first to confirm that typical gene regulatory networks can generate a close-to-optimal entangled distribution, and second to check whether Hill-type regulatory logic can combine correlated input noise, for example, add anticorrelated inputs, to maximize information transmission.

In the feed-forward motif, input TF X regulates TF Y , and both jointly regulate the expression level g of the target gene. Since each regulatory function can be either an activator (positive) or a repressor (negative), there are eight types of FFLs (16). If the sign of the direct regulation of g by X is the same as the sign of regulation of g by X through Y (sign of X -to- Y regulation multiplied by that of Y -to- g),

then the network is called coherent. The four networks that are not coherent are called incoherent. Moreover, if both X and Y are required to express g , the FFL is designated with an AND gate. If either TF can result in expression, an OR-gate designation is used. Including the gate, there are 16 unique types of FFL (Fig. 5). The model of competing TFs considered above closely resembles FFL Type 2 (+++OR; Fig. 5).

We have systematically simulated the stochastic dynamics of all FFL types for a range of input TF concentrations (see Methods). We focus on the intrinsic noise in the joint regulation of g by TFs X and Y , which has two sources: output noise due to stochastic synthesis and degradation of g , and input noise in readouts of TF concentrations. The input noise of TF X is correlated with that of Y with correlation coefficient ρ . The extrinsic contribution of noise from the upstream regulation of Y by X is considered negligible; its inclusion does not qualitatively change our results. The cross-regulation sets up the joint distribution of X and Y , $P(c_1, c_2)$, for a given input distribution of X , $P_{in}(c_1)$.

The probability of observing expression level g for input concentration c_1 , $P(g|c_1)$, is computed by sampling the steady-state expression levels for many runs. As is evident in Fig. 5 A, the noise distribution $P(g|c_1)$ is not Gaussian

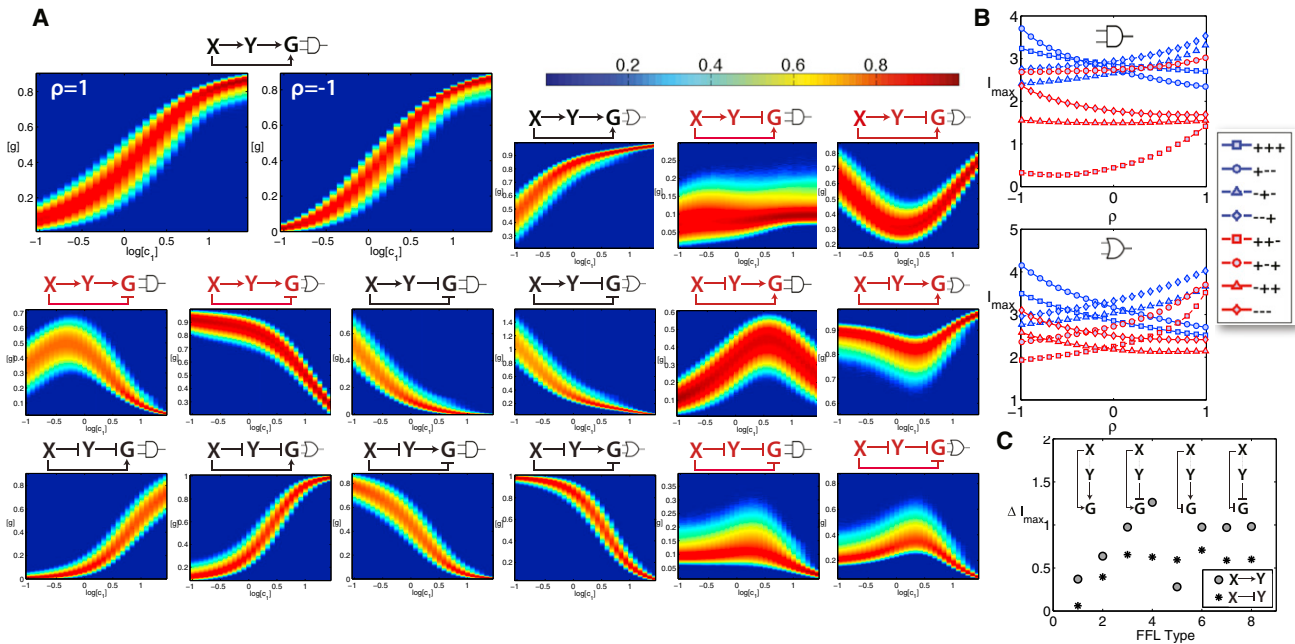


FIGURE 5 Numerical simulation of the feed-forward motif. There are 16 types of FFLs. The eight incoherent networks are denoted in red. (A) Probability of output expression level, g , as a function of log input X concentration c_1 , $p(g|c_1)$, normalized to the displayed color bar. Incoherent networks can exhibit nonmonotonic response. The larger figures show the response for Type 1 coherent FFL when $\rho = 1$ and $\rho = -1$, where ρ is the X - Y noise correlation coefficient. Anticorrelated input noise results in reduced uncertainty in response g and a higher channel capacity. The smaller figures have $\rho = 0$. (B) Channel capacity (bits) for all FFLs as a function of input-noise correlation coefficient ρ (AND-gate networks top, OR-gate networks bottom). The legend denotes the sign (+, activator; -, repressor) of X - Y , X - G , and Y - G regulation, respectively. The incoherent networks (red curves) have generally lower channel capacity than the coherent ones (blue curves). Networks where X and Y regulate G with the same sign (solid curves) have enhanced channel capacity for negative ρ , and those with opposite sign (dashed curves) improve with positive ρ . (C) For each network type, the circle (star) denotes maximum gain in capacity from noise correlations for X activating (repressing) Y . Odd and even network numbers correspond to AND and OR gates, respectively. The upstream cross-regulation between X and Y is important in taking advantage of noise correlations.

in general; nonlinearities in the model result in lopsided distributions. Furthermore, the incoherent FFLs are nonmonotonic functions of input c_1 to g . The same expression level g can correspond to more than one intended input, also in contrast to the earlier assumptions.

We computed the channel capacity of each FFL by numerically optimizing mutual information between output distribution, $p(g) = \int P(g|c_1)P_{in}(c_1)dc_1$, and TF X input distribution, $P_{in}(c_1)$ over all input distributions (see [Methods](#)). Coherent FFLs have generally a higher capacity than the incoherent ones; on average, coherent AND loops can transmit 2.8 bits vs. 1.6 bits for incoherent ANDs, whereas for OR networks the capacity is 3.1 bits vs. 2.4 bits. The nonmonotonicity of the incoherent networks creates ambiguities in mapping the output to the intended input. Correlations in fluctuations of readouts of X and Y concentrations increase channel capacity in all FFLs ([Fig. 5 B](#)). Networks where X and Y regulate g with the same sign (both activate or both repress g) enhance their channel capacity when the input noise correlation coefficient, ρ , is negative; networks with opposite signs improve with positive ρ . This is expected, since when X and Y regulate g in the same way, inputs are effectively added; adding two channels with anticorrelated noise reduces the noise. In a similar way, subtracting inputs with correlated noise results in noise reduction. For our choice of parameters, we observe, for example, a 13% increase in channel capacity of Type 1 Coherent FFL, +++AND (sign of X - Y , X - G , and Y - G regulation, respectively), when $\rho = -1$ compared to $\rho = 0$; coherent FFL +--OR, which roughly correspond to our earlier model of TFs competing for the same binding site, showed an improvement of 31%. Other choices of parameters produced similar results.

We claimed above that interacting TFs with an optimal-input joint distribution outperformed noninteracting TFs when noise correlations were incorporated, despite the increase in the noise of the individual channels from competition. This observation broadly holds in our simulations. For example, the highest-capacity network, FFL --+OR, has a channel capacity of 4.0 bits, with an input-noise correlation coefficient $\rho = 1$; whereas the same network with the variance of input noise reduced by a factor of 2 ($q = 0.5$; see [Methods](#)) and no correlation has a capacity of 3.7 bits. Incorporating correlations at the expense of higher noise variance seems to be a beneficial strategy. Lastly, we stressed the importance of up-stream cross-regulation between the two TFs as a means of constructing the optimal entangled joint distribution. To confirm that the gain in capacity from noise correlations is not simply due to a reduction in noise by a clever addition/subtraction of the inputs at the module regulating g , we compared the maximum gain in capacity from correlations for each network to its sister network (sign of X - Y regulation flipped). [Fig. 5 C](#) shows that the upstream X - Y regulation is instrumental in determining the gain; whether X activates or represses Y further

optimizes the input joint distribution and in turn the channel capacity.

DISCUSSION

We showed that quite generally, a signaling pathway with interference, that is, correlations in input noise due to microscopic interactions of the signaling molecules, can optimize information transmission by implementing cross talk upstream between the interacting molecules, such that the concentration of one input depends on the other.

Concentration-dependent transcriptional regulation is particularly important at the developmental stage. The concentrations of morphogens dictate cell fate, for example, resulting in patterning of the *Drosophila* embryo along the dorsoventral axis (38). It is likely that the embryo has optimized information transmission to ensure accurate patterning and later development. Gene regulation using a combination of TFs is also a common theme in development (39).

Xu et al. (40) have observed the feed-forward motif in regulation of the gene *Race* in the *Drosophila* embryo. They report that the intracellular protein Smads sets the expression level of *zerknüllt* (*zen*), and then Smads in combination with *zen* (twofold input) directly activates *Race*. Analysis of the binding site of Smads and *zen* reveals slight overlaps, and experiments indicate that one protein facilitates binding of the other to the enhancer. This interaction can result in a similar positive correlation coefficient in the TF-concentration estimates derived above. The previously proposed suggestion (40) that the feed-forward motif increases sensitivity to the input signal does not explain why the target is regulated by both the initial input and the target TF. Proposed dynamical features associated with the FFL (15,16) do not explain the need for overlapping binding sites and TF interactions at binding.

Another example of a feed-forward motif coupled to binding interactions is the joint regulation of *even-skipped* (*eve*) stripe 2 by *bicoid* (*bcd*) and *hunchback* (*hb*). Small et al. (3) report cooperative binding interactions between *bcd* and *hb* and a clustering of their binding sites in the promoter region. Upstream, *bcd* positively regulates transcription of *hb*. Similarly, Ip et al. (4) have observed joint activation of the gene *snail* (*sna*) by *twist* (*twi*) and *dorsal* (*dl*), which also exhibit cooperative binding interactions. *dl* directly regulates transcription of *twi* upstream.

More generally, other forms of cross talk besides transcriptional regulation can be used. For instance, in regulation of anaerobic respiration in *Escherichia coli*, regulators NarP and NarL are jointly regulated through phosphorylation by histidine kinase NarQ. NarL is also phosphorylated by kinase NarX. Downstream, NarP and NarL share the same DNA binding site (41). However, it is not clear whether optimizing channel capacity is relevant for this system. It is also possible that interference is

implemented using schemes other than DNA binding, for example, through cooperative interactions of signaling molecules with scaffold proteins (42).

We have shown that compared to noninteracting TFs, TFs interacting at overlapping binding sites and with upstream cross-regulation can enhance information transmission. This is consistent with frequent observations of the feed-forward motif ending in overlapping binding sites in developmental gene networks. Although it has been proposed previously that the feed-forward motif can optimize information transmission in regulatory networks (30), we emphasize that our approach is fundamentally different, since it stems from correlated binding noise, and it requires the physical existence of TF interactions at the binding level. This is indeed what is experimentally observed in the three examples discussed above. Diamond motifs, where inputs are transmitted independently and then recombined later, have also been proposed as mechanisms for increasing gain in signaling pathways (43).

The key assumptions in the above model were as follows. The primary source of noise is intrinsic input noise from readouts of TF concentrations, and extrinsic noise from cross-regulation is negligible. Inclusion of extrinsic noise simply reduces the overall channel capacity and does not modify the relation between input noise correlations and upstream cross-regulation. In the case of TFs competing for the same binding site, the intrinsic noise was assumed to be dominated by binding fluctuations as opposed to diffusion noise. This assumption is valid in the limit of low dissociation rates and low cooperativity (44). Moreover, this limit can be potentially consistent with biologically relevant integration times: competing TFs with low dissociation rates have a higher channel capacity than independent TFs, even when integration time is comparable to the typical mRNA lifetime; with negative cooperativity, the integration time can be arbitrarily large. Even if noise is dominated by diffusion, other mechanisms—such as di- or multimerization of the signaling molecules, or cooperative active transport—may generate correlations in diffusion noise. The same framework can then connect multimerization of signaling molecules to their upstream cross-regulation. Although our example focused on the particular case of competing TFs, we stress that in general any signaling pathway with correlated noise can transmit more information when optimized with cross talk between the inputs.

A myriad of logical regulatory circuits have been proposed through the use of overlapping binding sites and interacting TFs (26,32). It is worthwhile to see whether the upstream TF regulatory network of these systems can be correctly predicted from TF binding-site overlap or other interactions using the methodology outlined above (Eq. 6). Such analysis requires knowledge of the input noise, which can be obtained by a bioinformatics approach, where the binding sequence of each TF is examined for overlap, or

by direct measurement of the noise using single-molecule techniques (37) for other types of interactions.

The author thanks Boris Shraiman for helpful discussions and critical reading of the manuscript and Bill Bialek for introducing me to the subject.

This research was supported in part by the National Science Foundation under Grant No. NSF PHY11-25915.

REFERENCES

1. Wolpert, L. 2006. Principles of Development, 3rd ed. Oxford University Press, New York.
2. Wolpert, L. 1969. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* 25:1–47.
3. Small, S., A. Blair, and M. Levine. 1992. Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J.* 11:4047–4057.
4. Ip, Y. T., R. E. Park, ..., M. Levine. 1992. *dorsal*-twist interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev.* 6:1518–1530.
5. Elowitz, M. B., A. J. Levine, ..., P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science.* 297:1183–1186.
6. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA.* 99:12795–12800.
7. Paulsson, J. 2004. Summing up the noise in gene networks. *Nature.* 427:415–418.
8. Tkačik, G., and A. M. Walczak. 2011. Information transmission in genetic regulatory networks: a review. *J. Phys. Condens. Matter.* 23:153102.
9. Gregor, T., D. W. Tank, ..., W. Bialek. 2007. Probing the limits to positional information. *Cell.* 130:153–164.
10. Shannon, C. E. 1949. Communication in the presence of noise. *Proc. Inst. Radio Eng.* 37:10–21.
11. Cover, T. M., and J. A. Thomas. 1991. Elements of Information Theory. Wiley, New York.
12. Tkacik, G., C. G. Callan, Jr., and W. Bialek. 2008. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA.* 105:12265–12270.
13. Higham, D. J. 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev. Soc. Ind. Appl. Math.* 43:525–546.
14. Papoulis, A. 1991. Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York.
15. Shen-Orr, S. S., R. Milo, ..., U. Alon. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31:64–68.
16. Mangan, S., and U. Alon. 2003. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA.* 100:11980–11985.
17. Tyson, J. J., K. C. Chen, and B. Novak. 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15:221–231.
18. Gregor, T., W. Bialek, ..., E. F. Wieschaus. 2005. Diffusion and scaling during early embryonic pattern formation. *Proc. Natl. Acad. Sci. USA.* 102:18403–18407.
19. Powell, M. J. D. 1978. A fast algorithm for nonlinearly constrained optimization calculations. In Numerical Analysis. G. A. Watson, editor. Springer Verlag, Berlin. 144–157.
20. Tkacik, G., C. G. Callan, Jr., and W. Bialek. 2008. Information capacity of genetic regulatory elements. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 78:011910.

21. Little, S. C., G. Tkačik, ..., T. Gregor. 2011. The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. *PLoS Biol.* 9:e1000596.
22. Raser, J. M., and E. K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. *Science.* 304:1811–1814.
23. Newman, J. R., S. Ghaemmaghami, ..., J. S. Weissman. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature.* 441:840–846.
24. Rosenfeld, N., J. W. Young, ..., M. B. Elowitz. 2005. Gene regulation at the single-cell level. *Science.* 307:1962–1965.
25. Lee, T. I., N. J. Rinaldi, ..., R. A. Young. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science.* 298:799–804.
26. Hermesen, R., S. Tans, and P. R. ten Wolde. 2006. Transcriptional regulation by competing transcription factor modules. *PLOS Comput. Biol.* 2:e164.
27. Bialek, W., and S. Setayeshgar. 2005. Physical limits to biochemical signaling. *Proc. Natl. Acad. Sci. USA.* 102:10040–10045.
28. Kubo, R. 1966. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* 29:255–284.
29. Berg, H. C., and E. M. Purcell. 1977. Physics of chemoreception. *Biophys. J.* 20:193–219.
30. Walczak, A. M., G. Tkacik, and W. Bialek. 2010. Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 81:041905.
31. Tkačik, G., A. M. Walczak, and W. Bialek. 2012. Optimizing information flow in small genetic networks. III. A self-interacting gene. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 85:041903.
32. Buchler, N. E., U. Gerland, and T. Hwa. 2003. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA.* 100:5136–5141.
33. Bintu, L., N. E. Buchler, ..., R. Phillips. 2005. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* 15:125–135.
34. Shahrezaei, V., and P. S. Swain. 2008. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA.* 105:17256–17261.
35. Kaern, M., T. C. Elston, ..., J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6:451–464.
36. Golding, I., J. Paulsson, ..., E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell.* 123:1025–1036.
37. Wang, Y., L. Guo, ..., N. P. Ong. 2009. Quantitative transcription factor binding kinetics at the single-molecule level. *Biophys. J.* 96:609–620.
38. Morisato, D., and K. V. Anderson. 1995. Signaling pathways that establish the dorsal-ventral pattern of the *Drosophila* embryo. *Annu. Rev. Genet.* 29:371–399.
39. Howard, M. L., and E. H. Davidson. 2004. *cis*-Regulatory control circuits in development. *Dev. Biol.* 271:109–118.
40. Xu, M., N. Kirov, and C. Rushlow. 2005. Peak levels of BMP in the *Drosophila* embryo control target genes by a feed-forward mechanism. *Development.* 132:1637–1647.
41. Stewart, V. 2003. Biochemical Society Special Lecture. Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria. *Biochem. Soc. Trans.* 31:1–10.
42. Good, M. C., J. G. Zalatan, and W. A. Lim. 2011. Scaffold proteins: hubs for controlling the flow of cellular information. *Science.* 332:680–686.
43. de Ronde, W. H., F. Tostevin, and P. R. Ten Wolde. 2012. Feed-forward loops and diamond motifs lead to tunable transmission of information in the frequency domain. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 86:021913.
44. Tkacik, G., T. Gregor, and W. Bialek. 2008. The role of input noise in transcriptional regulation. *PLoS ONE.* 3:e2774.