Effective Transcription with Low-Affinity Binding Sites

Shengdao Ke[1] and Zifang Wu[2]

[1]Phillips Academy Andover, Andover, MA
[2]Bellevue High School, Bellevue, WA

October 17, 2026

Abstract

Multicellular eukaryotes often regulate transcription through clusters of low-affinity binding sites rather than single high-affinity sites. Such architectures may help cognate transcription factors distinguish functional regulatory regions from the many decoy sequences distributed throughout large genomes. However, clustering weak sites introduces a competing challenge: non-cognate transcription factors can also bind to the same regulatory region, reducing specificity through competition with the cognate factor. Here, we study these binding-site clusters using an information-theoretic framework, focusing on how binding-site number and affinity affect the reliability of information transmission from transcription-factor concentration to gene expression. Numerical optimization shows that information transmission depends jointly on site number and binding affinity. We find that at higher transcription-factor concentrations, promoters with multiple binding sites outperform single-site architectures, providing a possible information-theoretic explanation for the prevalence of clustered regulatory regions in eukaryotic genomes.

Keywords: transcription factors; low-affinity binding sites; information theory; gene regulation

Acknowledgements

We thank our mentors David Darrow and Joseph Landsittel for their invaluable guidance, feedback, and ideas throughout the research process. We extend our gratitude to the MIT-PRIMES-USA program for providing this research opportunity.

Contents

1 Introduction **4**
1.1 Mathematical framework of gene regulation 4
1.1.1 The Monod–Wyman–Changeux model 5
1.2 Noise in gene transcription 6
1.3 Information theory 7

2 Results **8**
2.1 Binding site number and response 8
2.2 One-gene information theoretic optimal promoter architecture 8
2.2.1 Limitations of this model 9
2.3 Promoter architecture in the presence of a non-cognate transcription factor 10
2.4 Beyond the single-gene model 11

3 Discussion **13**

A Derivation of noise in gene expression **14**

B Derivation of information-optimizing input distributions **15**

C Derivation of variance **16**
C.1 Variance of simple one-gene binding 16
C.2 Variance of the regulatory cascade motif 17

D Non-cognate transcription factor MWC model **18**

1 Introduction

To facilitate protein production, a transcription factor must recognize a short length of DNA, typically six to ten base pairs, and find that sequence within a genome much larger than it. The difficulty of this task depends on genome length; a bacterial genome is typically only a few million base pairs long, meaning that a given sequence only appears once and a single site possesses enough recognizability. However, the genome of a multicellular eukaryote can be substantially longer; that same sequence can thus appear multiple times. Binding to said sequence then does not necessarily identify a functional regulatory site.

This problem can be phrased in bits. Locating one unique location among N base pairs requires $\log_2 N$ bits under a simplified model; a typical multicellular eukaryote requires approximately 30 bits of information, while a bacterial genome only needs 22 [Wunderlich and Mirny, 2009]. Bacterial binding motifs carry 23 bits on average, which allows them to match a single site to its target. However, multicellular eukaryote motifs only carry around 12, making it difficult for them to correctly identify functional sites in the presence of $10^3 - 10^5$ accessible decoy sites on the genome [Wunderlich and Mirny, 2009]. To allow transcription factors to distinguish between binding sites, clustering of binding sites has been observed in literature [Wunderlich and Mirny, 2009].

Clusters of the same type of binding site often contain multiple sequences that are recognized by the same transcription factor; they also appear in over half of human promoters [Gotea et al., 2010]. Such clusters often include low-affinity binding sites. The binding affinity of a site is characterized by its dissociation constant K_d, which is measured by the transcription factor concentration needed to achieve 50% occupancy; low-affinity sites have larger K_d values and require higher concentrations. Research has shown that clusters of weak Zif268 binding sites within yeast promoters can produce substantial output; four to six weak sites were shown to produce expression comparable to three high-affinity consensus sites [Shahein et al., 2022]. Furthermore, research research has shown that high activity also appears to trade off with high specificity in human transcription factors [Naderi et al., 2024]. Although the low-affinity clusters may allow transcription factors to more effectively find its target promoter, it remains unclear how the ability of transcriptional output to reliably respond to changes in cognate transcription-factor concentration despite competition from non-cognate factors is retained.

To evaluate this, we view transcription factor concentration as the input to some regulatory system and gene expression as the system's output; inputting some fixed concentration can produce a distribution of possible expression levels. In the case where multiple distributions from different concentrations have significant overlap, the output cannot reliability distinguish between those input concentrations. Thus, we use mutual information to quantify how much observing gene expression can reduce the uncertainty about transcription factor concentration. Previous works have applied this approach to optimize transcriptional regulation [Tkacik et al., 2008a].

We address this question by deriving information theoretic equations relating binding site number, dissociation constant value, and transcription factor concentration. Through optimization, we find a direct relationship between the dissociation constant and the number of binding sites, corresponding to an inverse relation between binding affinity and binding site number. Furthermore, we find that a larger number of low-affinity sites is crucial to maintaining information transmission, especially in the presence of a non-cognate transcription factor.

1.1 Mathematical framework of gene regulation

We begin with the simple case of a single transcription factor and a single binding site. Let c denote the transcription factor concentration, n the mean occupancy of the binding site, $G = G(t)$ the current amount of protein, and R the transcription rate when the transcription factor is bound. Transcription factor binding is a stochastic process in which binding and unbinding occur with association rate $k_+ c$ and dissociation rate k_-, respectively. This process is illustrated in Figure 1

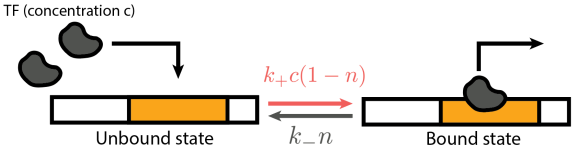


Figure 1: The binding site alters between two states: bound or unbound. The transition rates are given by $k_+c(1-n)$ and k_-n, where k_- and k_+ are the dissociation and association rates, respectively. For the bound state, $n = 1$ and for the unbound state, $n = 0$. We assume that transcription is active only in the bound state.

and the mean occupancy evolves according to the equation

$$\frac{dn}{dt} = k_+c(1-n) - k_-n. \tag{1.1}$$

To describe how the transcription factor concentration c regulates gene expression, we introduce the input-output function $f(c) \in [0,1]$, giving the activation of a gene. In the case of one binding site, $f(c)$ is determined by the mean binding site occupancy. Protein is produced at a maximum rate R and has lifetime τ. The dynamics of protein production is thus

$$\frac{dG}{dt} = -\frac{1}{\tau}G + Rf(c). \tag{1.2}$$

As protein proteins are synthesized and degraded, the system eventually reaches a steady-state, where the mean protein amount no longer changes over time and $dG/dt = 0$. At maximal expression, $f(c) = 1$, giving $G_{\max} = R\tau$. As it is useful to normalize the protein abundance, we define $N_{\max} = R\tau$ as the maximum protein abundance and introduce the dimensionless expression level $g := G/N_{\max}$. The protein dynamics can then be written as

$$\tau\frac{dg}{dt} = f(c) - g. \tag{1.3}$$

We use the unnormalized protein abundance G and the normalized expression level g throughout the paper. For the binding process, we assume that transcription factor association and dissociation reach an equilibrium at a fixed concentration c, where $k_+c(1-n) - k_-n_{eq}$, giving mean binding site occupancy

$$n_{\text{eq}} = \frac{c}{c + K_d}, \tag{1.4}$$

where $K_d = k_-/k_+$ is the dissociation constant. In particular, K_d is precisely the concentration at which the site is occupied half of the time, on average.

For systems in which binding is cooperative or the regulatory response is more switch-like, a single binding site can be modeled (phenomenologically) as a Hill function [Hill, 1910]

$$f(c) = \frac{c^h}{K_d^h + c^h}, \tag{1.5}$$

where h is the Hill coefficient, which controls the steepness of the response and K_d is the dissociation constant.

These dynamics can be generalized to systems of multiple genes $\{g_i\}$ by introducing $f(c; \{g_j\}) \in [0,1]$, the function describing the activation rate of gene i given the transcription factor concentration c and expression levels of other genes g_j. The dynamics of a gene i can then be described as

$$\tau\frac{dg_i}{dt} = f_i(c; \{g_j\}) - g_i. \tag{1.6}$$

Equivalently, under the unnormalized protein abundance G, the dynamics become

$$\frac{dG_i}{dt} = Rf_i\left(c; \left\{\frac{G_j}{N_{\max}}\right\}\right) - \frac{G_i}{\tau}. \tag{1.7}$$

With these multi-gene systems, we describe $f(c; g_i)$ through the Monod–Wyman–Changeux model, introduced below.

1.1.1 The Monod–Wyman–Changeux model

The Monod–Wyman–Changeux (MWC) model was developed for describing cooperative allosteric transitions (Monod et al. [1965]). In our application, the promoter is assumed to be in one of two states: a transcriptionally inactive "off" state and a transcriptionally active "on" state. Rather than

assigning the promoter deterministically to one of these states, statistical mechanics associates each state with a statistical weight determined by its free energy. The probability of occupying a given state is then equal to that state's statistical weight divided by the sum of the weights of all accessible states [Phillips and Orme, 2020]. We denote the dimensionless free-energy difference between the inactive and active promoter states by $\Delta\varepsilon$, measured in units of $k_{\mathrm{B}}T$. We assume that the number of binding sites n stays consistent between the "on" and "off" states.

Denoting the corresponding dissociation constants by $K_{d,\mathrm{on}}$ and $K_{d,\mathrm{off}}$, the statistical weights of each binding the two promoter states are

$$Z_{\mathrm{on}} = \left(1 + \frac{c}{K_{d,\mathrm{on}}}\right)^{n}, \tag{1.8}$$

and

$$Z_{\mathrm{off}} = e^{\Delta\varepsilon}\left(1 + \frac{c}{K_{d,\mathrm{off}}}\right)^{n}. \tag{1.9}$$

Furthermore, as in Tkačik and Walczak [2011], we assume the limit $c/K_{d,\mathrm{off}} \ll 1$, so (1.1.1) simply becomes $Z_{\mathrm{on}} = e^{\Delta\varepsilon}$.

The total promoter partition function is the sum of the statistical weights of the two promoter states [Phillips and Orme, 2020],

$$Z = Z_{\mathrm{off}} + Z_{\mathrm{on}} = e^{\Delta\varepsilon} + \left(1 + \frac{c}{K_d}\right)^{n}. \tag{1.10}$$

The probability that the promoter is in the active state is therefore

$$\begin{aligned} f(c) &= \frac{Z_{\mathrm{on}}}{Z_{\mathrm{off}} + Z_{\mathrm{on}}} \\ &= \frac{1}{1 + e^{\Delta\varepsilon}\left(1 + \frac{c}{K_d}\right)^{-n}}. \end{aligned} \tag{1.11}$$

Throughout this work, we identify $f(c) = P(\mathrm{on} \mid c)$ with the activation function. This framework becomes especially useful when dealing with more complicated systems of multiple binding sites, genes, and transcription factors.

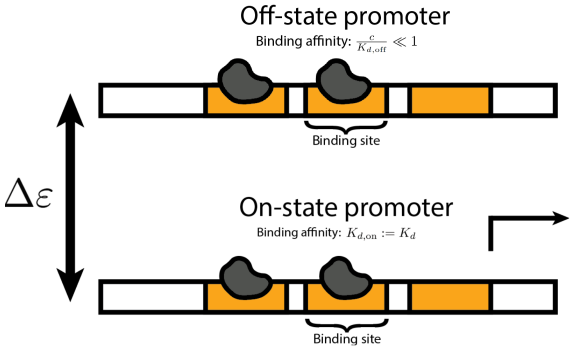


Figure 2: Reproduced MWC promoter model from Tkačik and Walczak [2011]. The promoter can occupy either an off state or an on state, separated by an intrinsic energetic bias $\Delta\varepsilon$. We assume $c/K_{d,\mathrm{off}} \ll 1$, so cognate binding contributes negligibly to the off-state statistical weight, while binding to the on state remains appreciable.

1.2 Noise in gene transcription

Gene expression is inherently stochastic. Even with the same regulatory environment, two identical cells may produce different amounts of proteins because the molecular events in regulatory processes are probabilistic. Transcription factors diffuse through the cell and randomly bind and unbind to binding regions, while produced proteins and mRNA stochastically degrade.

Consequently, a fixed transcription factor concentration c does not produce a single deterministic gene-expression level but rather a distribution $P(g \mid c)$ of possible outputs with mean expression $\bar{g}(c)$ and variance $\sigma_g^2(c) = \langle (g - \bar{g}(c))^2 \rangle$. With this in mind, we consider three primary biological sources of noise (Tkačik and Walczak [2011], Walczak et al. [2010]).

First, production and degradation noise arises from the discrete synthesis and removal of mRNA and proteins. Transcription, translation, and degradation occur as individual molecular events, so the amount of gene product fluctuates even if regulatory occupancy were held constant. For a simple birth-death process, these fluctuations are particularly important when the number of molecules is small. Second, transcription factors arrive at their target volume through stochastic

diffusion, which accounts for local concentration fluctuations even with a constant average cellular concentration. Third, if produced proteins act as transcription factors in these networks, they also contribute to noise in the same way as any other transcription factor.

These stochastic effects can be incorporated into otherwise deterministic regulatory equations using a Langevin white noise $\xi(t)$. For the rest of the paper, we include a noise term in the generalized gene dynamics (1.1) and write

$$\tau \frac{dg_i}{dt} = f_i(c;\{g_j\}) - g_i + \xi_i(t). \tag{1.12}$$

The Langevin noise terms $\xi_i(t)$ in the equation satisfy the properties

$$\langle \xi_i(t) \rangle = 0 \tag{1.13}$$

$$\langle \xi_i(t)\xi_j(t') \rangle = N_{ij}\delta(t-t'), \tag{1.14}$$

where δ is the Dirac delta function. The second property defines N_{ij}, the spectral density of the white Langevin noise. Thus, the coefficients N_{ij} form the noise spectral density matrix N, which characterizes the strength and correlations of the Langevin noise.

It is important to distinguish the noise matrix N from the covariance matrix C of the gene expression fluctuations. Defining $\delta g_i = g_i - \bar{g}_i$, the covariance matrix is

$$C_{ij} = \langle \delta g_i(t) \delta g_j(t) \rangle , \tag{1.15}$$

and describes the resulting correlations between fluctuations in the expression levels of genes i and j. In contrast, N_{ij} characterizes the strength and correlations of the stochastic Langevin forces $\xi_i(t)$ that drive these fluctuations.

To write the diffusion-noise terms while retaining physical concentrations, the model warrants a natural concentration scale

$$c_0 = \frac{N_{\max}}{D\ell\tau}, \tag{1.16}$$

where D is the diffusion constant and ℓ is the characteristic diffusion length [Walczak et al., 2010, Tkačik et al., 2009, Tkačik and Walczak, 2011]. If the product of gene k acts as a transcription factor, we let $c_{k,\max} = N_{\max}/\Omega_k$ denote the physical concentration corresponding to the normalized level $g_k = 1$, where Ω_k is the effective volume for that regulator. Assuming that the Langevin noise sources acting directly on different genes are independent, the noise matrix is diagonal, with

$$N_{ij} = \delta_{ij} \frac{\tau}{N_{\max}} \left[2\bar{g}_i + c_0 c \left(\frac{\partial f_i}{\partial c} \right)^2 + \sum_k \frac{c_0}{c_{k,\max}} \left(\frac{\partial f_i}{\partial g_k} \right)^2 \bar{g}_k \right]. \tag{1.17}$$

Thus, c_0 sets the natural concentration scale for diffusion noise. Equivalently, one may define the dimensionless variables $x = c/c_0$ and $x_{k,\max} = c_{k,\max}/c_0$, in which case the two diffusion terms become $x(\partial f_i/\partial x)^2$ and $x_{k,\max}^{-1}(\partial f_i/\partial g_k)^2 \bar{g}_k$, respectively. The complete derivation is provided in Appendix A.

1.3 Information theory

The preceding models describe how transcription factor concentration is converted into gene expression and how stochastic fluctuations introduce variability into this response. However, characterizing the mean response $\bar{g}(c)$ and its noise separately does not directly reveal how effectively a regulatory system can distinguish between different input concentrations. A useful way to quantify this ability is to treat gene regulation as a problem of information transmission, which is limited by stochastic noise.

Throughout the main text, c denotes the physical transcription-factor concentration. The natural scale $c_0 = N_{\max}/(D\ell\tau)$ allows us to define the dimensionless concentration $x = c/c_0$. Because x is a one-to-one rescaling of c, it does not change the mutual information: $I(c;g) = I(x;g)$. We therefore retain physical concentrations in the information-theoretic derivations, while c_0 enters explicitly through the diffusion-noise strength. The maximum accessible input concentration is denoted $c_{\max}$, with dimensionless range $x_{\max} = c_{\max}/c_0$.

In this framework, the transcription factor concentration c is regarded as the input to a regulatory channel, while the resulting gene-expression

level g is the output. If the output distributions corresponding to different concentrations overlap strongly, then the regulatory system cannot reliably distinguish those inputs. Conversely, if changes in c produce sufficiently distinct output distributions, more information about the input concentration is preserved in the gene-expression response.

For continuous variables, the uncertainty in the output is quantified by the Shannon differential entropy

$$H(g) = -\int dg\, P(g) \log_2 P(g). \tag{1.18}$$

The average uncertainty that remains in the output when the input is known is the conditional entropy

$$H(g \mid c) = -\int dc\, P(c) \int dg\, P(g \mid c) \log_2 P(g \mid c). \tag{1.19}$$

The mutual information between transcription-factor concentration and gene expression is then

$$\begin{aligned} I(c;g) &= H(g) - H(g \mid c) \\ &= \int dc \int dg\, P(c,g) \log_2 \left(\frac{P(c,g)}{P(c)P(g)} \right). \end{aligned} \tag{1.20}$$

Thus, $I(c;g)$ measures how much observing one variable reduces uncertainty about the other. Mutual information is symmetric, so $I(c;g) = I(g;c)$.

2 Results

In the following derivations, we study the impact of the number of binding sites and binding affinity on information transmission, specifically from transcription factor concentrations to gene expression. We first optimize a model containing one gene regulated by a single cognate transcription factor, establishing how binding site number and affinity influence information transmission in the simplest setting. We then extend this framework to a single gene exposed to both cognate and non-cognate transcription factors, allowing us to examine how competitive binding affects regulatory precision. Next, we consider a two-stage regulatory cascade, in which the product of one gene regulates a downstream gene and noise can propagate through the network. Finally, we discuss how these ideas generalize to larger genetic networks containing multiple interacting genes and regulatory motifs.

2.1 Binding site number and response

In this section, we derive and analyze the dynamics of genes with multiple binding sites. Suppose we have a single gene with n binding sites and that all sites have an equal dissociation constant K_d. We assume $c/K_{d,\text{off}} \ll 1$, so cognate transcription-factor binding to the inactive promoter state contributes negligibly to its statistical weight.

Let $f(c) \in [0,1]$ be the probability that the promoter is active at some concentration of transcription factor c; note that the mean production rate is then $R(f(c))$, where R is the maximum production rate. At steady state, the mean expression $\bar{g}(c)$ is then

$$\bar{g}(c) = f(c) = \frac{1}{1 + e^{\Delta\varepsilon} \left(1 + \dfrac{c}{K_d}\right)^{-n}}. \tag{2.1}$$

where $\Delta\varepsilon$ is the dimensionless energetic bias toward the inactive promoter.

At fixed K_d and $\Delta\varepsilon$, increasing the number of binding sites increases the mean expression. For $\Delta\varepsilon > 0$, the concentration at which $f(c) = \frac{1}{2}$ is:

$$c_{1/2} = K_d \left(e^{\Delta\varepsilon/n} - 1 \right).$$

It follows that a larger n leads to a lower $c_{1/2}$. Furthermore, when binding weakens and K_d increases, $c_{1/2}$ increases. Also note that at larger c values, further increases lead to relatively smaller changes in expression. As such, we examine both the mean response and the variability in the following sections.

2.2 One-gene information theoretic optimal promoter architecture

Maximal information transmission is the largest mutual information that a regulatory system can achieve between its input and output when the input distribution is chosen optimally. Maximizing information transmission allows cells to respond more accurately to changes in their environment

and distinguish between signals even in the presence of noise (Tkačik et al. [2009], Ziv et al. [2007]). We first examine a single gene, potentially with many binding sites, regulated by one transcription factor. Because c is retained as a physical concentration, the single-gene noise spectral density, as defined by (1.2) is

$$N(c) = \frac{\tau}{N_{\max}}\left[2\bar{g}(c) + c_0 c\left(\frac{df}{dc}\right)^2\right], \tag{2.2}$$

which is the sum of synthesis/degradation noise and transcription-factor diffusion noise. The factor c_0 converts the physical concentration into the natural diffusion-noise scale.

Past studies (Tkačik et al. [2008, 2009]) have shown that, in the small-noise limit $\sigma_g(c)/\bar{g}(c) \ll 1$, the information-optimizing input distribution is inversely related to the effective input uncertainty. In particular,

$$P^*(c) = \frac{1}{Z}\frac{1}{\sigma_c(c)}, \qquad Z = \int_0^{c_{\max}} dc\, \frac{1}{\sigma_c(c)}. \tag{2.3}$$

We provide a derivation in Appendix B. The corresponding maximal information transmission is

$$I^*(c;g) = \log_2\left(\frac{Z}{\sqrt{2\pi e}}\right). \tag{2.4}$$

To determine $\sigma_c(c)$, we convert output noise into an equivalent uncertainty in the physical input concentration. For small fluctuations around a given c, the mean response can be linearized as $\delta g \approx (d\bar{g}/dc)\delta c$ [Tkačik and Walczak, 2011]. Therefore,

$$\sigma_c(c) = \frac{\sigma_g(c)}{|d\bar{g}/dc|}. \tag{2.5}$$

Writing $g(t) = \bar{g} + \delta g(t)$, the linearized one-gene Langevin dynamics are

$$\tau\frac{d\delta g}{dt} = -\delta g + \xi(t). \tag{2.6}$$

In Appendix C.1, we derive the expression

$$\sigma_g^2(c) = \frac{N(c)}{2\tau}, \tag{2.7}$$

Combining (2.2) with (2.2)

$$I^*(c;g) = \log_2\left[\frac{1}{\sqrt{2\pi e}}\int_0^{c_{\max}} dc\sqrt{\frac{2\tau}{N(c)}}\left|\frac{d\bar{g}}{dc}\right|\right]. \tag{2.8}$$

At steady state, $\bar{g}(c)$ is given by (2.1).

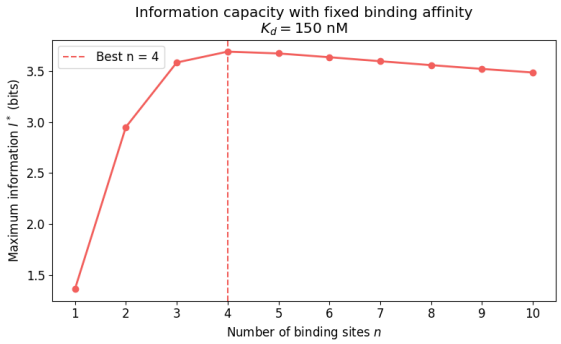
Having established an expression for maximal information transmission, we ask how the number of binding sites affects information transmission. The parameters n and K_d determine the shape of $\bar{g}(c)$, while c_0 sets the diffusion-noise concentration scale. In all numerical comparisons, c_0 is held fixed while n, K_d, and the accessible physical input range $0 \le c \le c_{\max}$ are varied. Equivalently, changing $c_{\max}$ at fixed c_0 changes the dimensionless range $x_{\max} = c_{\max}/c_0$. We therefore evaluate I^* over varying n and K_d. Figure 3 summarizes these results for both varying K_d and fixed K_d.

We find a direct relationship between the optimal dissociation constant and the number of binding sites in Figure 4, which is supported by experimental data from Shahein et al. [2022]. This suggests that optimizing information transmission is an important consideration in regulatory region architecture.

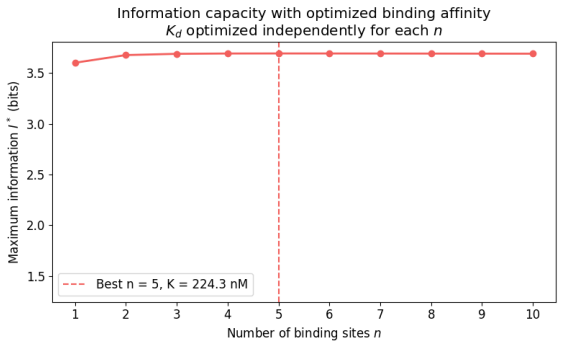
2.2.1 Limitations of this model

In Section 2.1, we made the assumption that all binding sites in a regulatory region have the same binding affinities. In biological systems, however, binding site sequences often differ, leading to site-specific dissociation constants [Shahein et al., 2022]. Accordingly, the dissociation constant K_d used in our model should be interpreted as an effective binding affinity rather than the microscopic dissociation constant of each individual site. We use this effective K_d as a way to summarize the combined effect of underlying dissociation constants rather than modeling each site individually.

Further, this effective K_d should not necessarily be directly compared with the cluster-level K_d reported by Shahein et al. [2022], which is obtained by fitting the total cluster occupancy to a single saturation curve. The two quantities are comparable only as they summarize the concentration scale over which the cluster responds.



(a)



(b)

Figure 3: (a) At a fixed binding affinity $K_d = 150\,\text{nM}$, promoters with more binding sites transmitted more information than the single-site architecture, with the optimal number being $n = 4$. This shows that at relatively high dissociation constants, architectures of more than one binding site are advantageous. (b) Allowing the dissociation constant to be varied freely, the maximal information transmission stays relatively flat across the number of binding sites. Compared to (a), this shows that different binding-site numbers require different dissociation constants to maximize information transmission. The optimal number in this case was $n = 5$. For both graphs, the parameters used were $c_0 = 100\,\text{nM}$ [Tkačik and Walczak, 2011], $N_{\text{max}} = 1000$, $c_{\text{max}} = 1000\,\text{nM}$, and $\Delta\varepsilon = \ln 99$, corresponding to a basal expression level of 1% in the absence of transcription factors.

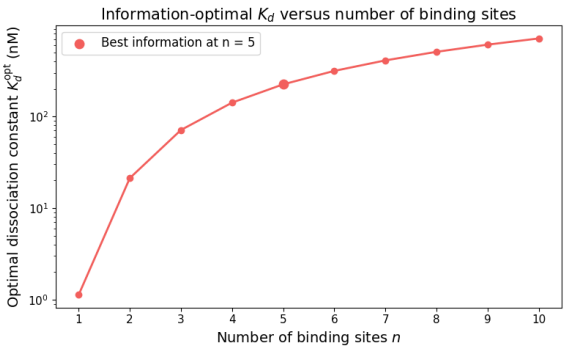


Figure 4: The information optimal dissociation constant increases as the binding site number increases. Thus, promoter architectures with fewer binding sites require stronger binding affinity to transmit information effectively. Maximizing information transmission serves as one reason why promoter regions with more binding sites tend to have lower binding affinities (higher dissociation constants), as found in Shahein et al. [2022].

2.3 Promoter architecture in the presence of a non-cognate transcription factor

In the previous section, we explored gene promoter regions with multiple binding sites under the assumption that only one cognate transcription factor could bind. In a cellular environment, however, non-cognate transcription factors can also bind weakly to the same sites and compete with the cognate factor [Zabet and Adryan, 2013]. We therefore extend the model to include a non-cognate transcription factor with physical concentration c_{nc} and dissociation constant K_{nc}. Thus, each site is assumed to be either unbound, cognate-bound, or non-cognate-bound.

The competitive MWC model derived in Appendix D can be written as

$$\begin{aligned} f(c, c_{\text{nc}}) &= \frac{1}{1 + e^{F(c, c_{\text{nc}})}}, \\ F(c, c_{\text{nc}}) &= \Delta\varepsilon + n \ln\left(1 + \frac{c_{\text{nc}}}{K_{\text{nc}}}\right) \\ &\quad - n \ln\left(1 + \frac{c}{K_d} + \frac{c_{\text{nc}}}{K_{\text{nc}}}\right). \end{aligned} \tag{2.9}$$

For the information calculation, we distinguish the fixed mean non-cognate background from its microscopic fluctuations by writing $c_{\text{nc}}(t) = \bar{c}_{\text{nc}} + \delta c_{\text{nc}}(t)$ and $\langle \delta c_{\text{nc}} \rangle = 0$. The cognate concentration c is the information-bearing input, whereas $\bar{c}_{\text{nc}}$ is held fixed within each capacity calculation.

Nevertheless, fluctuations in the non-cognate population still contribute to the output noise. Retaining physical concentrations, the noise spectral density is

$$N_{\rm nc}(c;\bar{c}_{\rm nc}) = \frac{\tau}{N_{\max}}\left[2\bar{g}(c;\bar{c}_{\rm nc}) + c_0 c\left(\frac{\partial f}{\partial c}\right)^2 + c_0\bar{c}_{\rm nc}\left(\frac{\partial f}{\partial c_{\rm nc}}\right)^2\right], \tag{2.10}$$

where the derivatives are evaluated at $c_{\rm nc} = \bar{c}_{\rm nc}$. The final term is the additional diffusion noise generated by the fluctuating non-cognate population.

Under the same small-noise approximation used above,

$$\sigma_g^2(c;\bar{c}_{\rm nc}) = \frac{N_{\rm nc}(c;\bar{c}_{\rm nc})}{2\tau}, \quad \sigma_c(c;\bar{c}_{\rm nc}) = \frac{\sigma_g(c;\bar{c}_{\rm nc})}{|\partial\bar{g}/\partial c|}. \tag{2.11}$$

Because $\bar{c}_{\rm nc}$ is a fixed background parameter rather than a second information bearing input, the optimization remains one-dimensional. From Appendix D, we thus still have

$$P^*(c \mid \bar{c}_{\rm nc}) = \frac{1}{Z(\bar{c}_{\rm nc})}\frac{1}{\sigma_c(c;\bar{c}_{\rm nc})}, \tag{2.12}$$

with

$$Z(\bar{c}_{\rm nc}) = \int_0^{c_{\max}} dc\,\frac{1}{\sigma_c(c;\bar{c}_{\rm nc})}. \tag{2.13}$$

The resulting conditional maximal information transmission is

$$I^*(c;g \mid \bar{c}_{\rm nc}) = \log_2\left[\frac{1}{\sqrt{2\pi e}}\int_0^{c_{\max}} dc\,\frac{|\partial\bar{g}/\partial c|}{\sigma_g(c;\bar{c}_{\rm nc})}\right]. \tag{2.14}$$

We next ask how non-cognate competition changes the architecture that maximizes information transmission. Both n and K_d determine the sensitivity of the promoter to the cognate input. For each value of $c_{\max}$, we set the mean background to $\bar{c}_{\rm nc} = \alpha c_{\max}$ and hold it fixed while c varies from 0 to $c_{\max}$. The results are summarized in Figure 5.

Compared with the single-transcription-factor model, the presence of non-cognate competition shifts the information-optimal promoter toward a higher dissociation constant. A larger K_d, corresponding to weaker individual binding, helps prevent saturation and maintains promoter sensitivity over a broader range of cognate transcription-factor concentrations. Increasing the number of binding sites compensates for this weaker affinity, allowing the promoter to preserve information transmission despite competition.

2.4 Beyond the single-gene model

In living organisms, genes rarely function as isolated input-output systems. Rather, they are often a part of a larger regulatory network, with interactions between genes forming recurrent regulatory motifs such as feedback and feed-forward loops that alter the dynamics of the system [Alon, 2007, Novák and Tyson, 2008]. As the number of genes increases, the dynamics and information transmission become increasingly difficult to model. Here, we provide a detailed analysis of the two-stage regulatory cascade, a motif of two genes where an external transcription factor activates the upstream gene, which in turn activates the downstream gene, as shown visually in Figure 6.

The two-stage cascade obeys

$$\begin{aligned}\tau\frac{dg_1}{dt} &= f_1(c) - g_1 + \xi_1(t),\\ \tau\frac{dg_2}{dt} &= f_2(g_1) - g_2 + \xi_2(t).\end{aligned} \tag{2.15}$$

The external input c is a physical concentration. The upstream output g_1, however, is dimensionless, so we introduce $c_{1,\max}$ as the physical concentration of gene-1 product corresponding to $g_1 = 1$. Thus, the physical concentration of the regulator acting on the downstream gene is $c_1 = c_{1,\max}g_1$. We use physical dissociation constants K_1 and K_2, giving the MWC activation functions

$$\begin{aligned}f_1(c) &= \frac{1}{1 + e^{\Delta\varepsilon_1}\left(1 + \dfrac{c}{K_1}\right)^{-n_1}},\\ f_2(g_1) &= \frac{1}{1 + e^{\Delta\varepsilon_2}\left(1 + \dfrac{c_{1,\max}g_1}{K_2}\right)^{-n_2}}.\end{aligned} \tag{2.16}$$

This distinction allows both K_1 and K_2 to be reported in physical units such as nM.

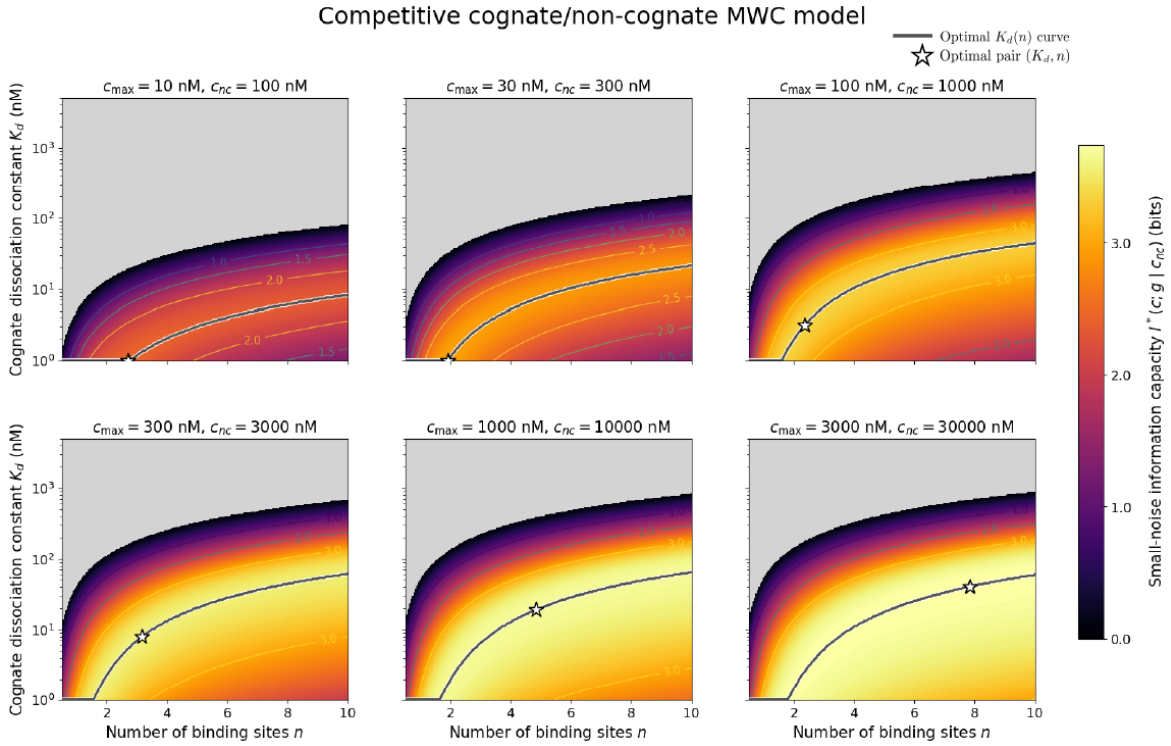


Figure 5: Similar to the one transcription factor case, the optimal dissociation constant K_d increases as the number of binding sites increases. Furthermore, we find that as the concentration of transcription factors increases, the optimal dissociation constant and binding site number also increase. This gives insight into the disparity discussed earlier between promoter region architecture in eukaryotes and prokaryotes, as eukaryotic cellular environments have comparatively higher numbers of transcription factors [Wunderlich and Mirny, 2009, Ma et al., 2005, Kao-Huang et al., 1977]. The gray regions indicate parameter values for which the small-noise approximation predicts a negative capacity and is therefore not quantitatively valid. The parameters used were $c_0 = 100\,\text{nM}$, $N_{\max} = 1000$, and $\Delta\varepsilon = \ln 99$.

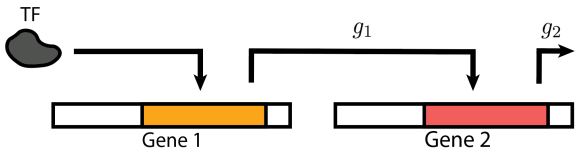


Figure 6: Visual representation of the regulatory cascade. A transcription factor regulates the production of g_1, which regulates the production of g_2.

Under the small-noise approximation, we linearize the regulatory functions around their steady-state values. The resulting fluctuation equations are linear and driven by approximately Gaussian Langevin noise (see Appendix B), so the same one-dimensional optimization over the external input applies, with $P^*(c) \propto 1/\sigma_c(c)$ and $I^* = \log_2(Z/\sqrt{2\pi e})$.

Assuming that the Langevin noise sources acting directly on the two genes are independent, the noise matrix is diagonal. Defining $x_{1,\max} = c_{1,\max}/c_0$, we obtain the two noise spectral densities

$$N_1 = \frac{\tau}{N_{\max}}\left[2\bar{g}_1(c) + c_0 c\left(\frac{\partial f_1}{\partial c}\right)^2\right],$$
$$N_2 = \frac{\tau}{N_{\max}}\left[2\bar{g}_2(c) + \frac{1}{x_{1,\max}}\left(\frac{\partial f_2}{\partial g_1}\right)^2 \bar{g}_1(c)\right]. \tag{2.17}$$

The diffusion term in N_1 is driven by the external transcription factor, whereas the diffusion term in N_2 is driven by the upstream gene's product. The scale $c_{1,\max}$ is therefore distinct from the external input limit $c_{\max}$, even if the two are assigned the same numerical value in a particular simulation. In the numerical cascade calculations reported below, we use $c_0 = 100\,\text{nM}$, $c_{\max} = 1000\,\text{nM}$, and $c_{1,\max} = 1000\,\text{nM}$, so that $x_{1,\max} = 10$.

Since g_1 is regulated directly by c, $I(c; g_1)$ is equivalent to the single-gene channel considered in Section 2.2. We therefore focus on the downstream information $I(c; g_2)$. In Appendix C.2, we show that in this case,

$$\sigma_c^2(c) = \frac{\sigma_{g_2}^2(c)}{\left(d\bar{g}_2/dc\right)^2} = \frac{2N_2 + N_1\phi_{21}^2}{4\tau\left(d\bar{g}_2/dc\right)^2}, \tag{2.18}$$

where

$$\phi_{21} = \left.\frac{\partial f_2}{\partial g_1}\right|_{\bar{g}_1(c)}. \tag{2.19}$$

We now analyze how the binding site number influences information transmission and the optimal dissociation constants. To reduce the number of parameters, we set $n_1 = n_2 = n$ while allowing K_1 and K_2 to vary independently. The calculated optimal dissociation constants are reported in Figure 8, and the corresponding channel capacities are summarized in Figure 7.

We find that to maximize information transmission, the dissociation constant K_2 of the downstream gene must be greater than that of K_1. It is important to make the distinction that the dissociation constant K_2 characterizes the binding affinity of the produce g_1 to binding sites on the downstream gene rather than the transcription factor dissociation constants we discussed previously. Figure 8 supports our previous finding that higher numbers of binding sites require higher dissociation constants to maintain information flow. Moreover, we find that in general, the information transmission from c to g_2 is lower than that of a single gene system, suggesting that the presence of noise from an intermediate channel gene 1 hinders information transmission.

In real cells, genetic networks are complicated systems often consisting of more than two genes.

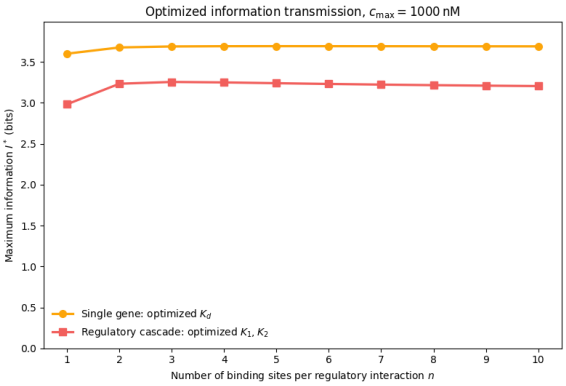


Figure 7: We find that under optimized dissociation constants, the maximal information transmission from a transcription factor c to one gene is consistently higher than when an intermediate gene is involved. This suggests that propagated noise from the upstream gene hinders the ability of the downstream gene to receive information. For each n, we set $n_1 = n_2 = n$ and optimize dissociation constants K_1 and K_2 independently from one another. In both cases, promoters of multiple binding sites transmit information more effectively than a single site. We use the parameters $c_0 = 100\,\text{nM}$, $c_{\max} = 1000\,\text{nM}$, and $c_{1,\max} = 1000\,\text{nM}$, so that $x_{1,\max} = 10$.

The resulting combinations of regulatory interactions, feedback loops, and stochastic fluctuations produce dynamics that are difficult to characterize analytically. Consequently, the principles governing information transmission through large gene-regulatory networks remain incompletely understood.

3 Discussion

Our analysis indicates that binding site number and binding affinity are inversely related under optimal information transmission constraints. Section 2.3 illustrates that in the presence of competition from non-cognate transcription factors, both optimal binding site number and optimal dissociation constant increase, corresponding to a decreased binding affinity. We note that our model is heavily dependent on the parameter values that we choose, and thus we are unable to recover a general optimal binding site number and dissociation

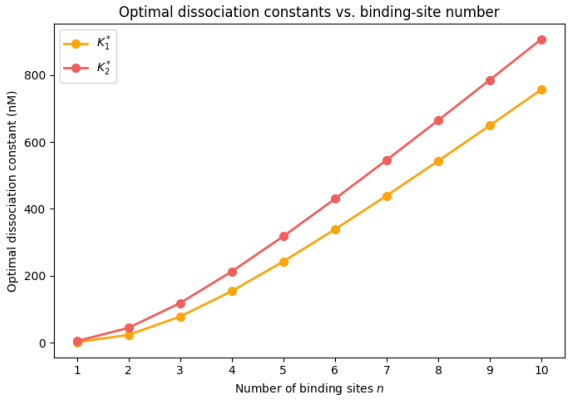


Figure 8: In the regulatory cascade, the information-optimal dissociation constant of the downstream is consistently higher than that of the upstream gene. There is a direct relationship between the number of binding sites and dissociation constant, as consistent with the one-gene case. The parameter values used were the same as in Figure 7.

constant. Nevertheless, the finding that optimal information transmission is achieved under multiple low-affinity binding sites is consistent across all models used. As reliable information transmission is fundamental to cellular function [Tkačik and Walczak, 2011], our observations above offer a possible information-theoretic explanation for the prevalence of low-affinity binding sites in eukaryotic regulatory regions Crocker et al. [2016], Shahein et al. [2022]. From this perspective, clusters of weak sites may represent a compromise between making a regulatory region distinguishable from genomic decoy sequences while maintaining sufficient specificity against competing non-cognate transcription factors.

Furthermore, our methods can be adapted to more complicated regulatory motifs while yielding similar results, as demonstrated in Section 2.4. We find that the presence of intermediate genes in transcriptional networks may hinder information transmission. Therefore, it remains a mystery how complicated regulatory architectures maintain information transmission.

Several simplifications of the present model leave room for future exploration. In particular, we assume an equilibrium MWC model and identical dissociation constants within a binding cluster. Real regulatory regions may instead contain cooperative or competitive interactions between binding sites [Crocker et al., 2015], and gene regulation can contain non-equilibrium binding [Mahdavi et al., 2024]. Incorporating these features may help determine whether the advantages of low-affinity binding clusters persist under more realistic conditions.

A Derivation of noise in gene expression

Recall the three sources of noise introduced in Section 1.2. To derive their contributions, it is useful to return to the unnormalized protein dynamics, where synthesis and degradation rates are expressed in molecules per unit time. We begin with

$$\frac{dG_i}{dt} = Rf_i\left(c; \left\{\frac{G_j}{N_{\max}}\right\}\right) - \frac{G_i}{\tau} + \Xi_i(t), \quad \text{(A.1)}$$

where $\Xi_i = N_{\max}\xi_i/\tau$ is the Langevin force in the unnormalized dynamics.

The first source, $\mathcal{N}_i^{(1)}$, is synthesis/degradation noise. For a birth–death process composed of single-molecule Poisson events, the noise spectral density is the sum of the mean synthesis and degradation rates [Walczak et al., 2010]. At steady state, $Rf_i = G_i/\tau$, so

$$\mathcal{N}_i^{(1)} = Rf_i + \frac{G_i}{\tau} = \frac{2G_i}{\tau}. \quad \text{(A.2)}$$

The second source is diffusion of the external transcription factor. Following Tkacik et al. [2008b] and Walczak et al. [2010],

$$\mathcal{N}_i^{(2)} = \left(R\frac{\partial f_i}{\partial c}\right)^2 \frac{c}{D\ell}. \quad \text{(A.3)}$$

Here D is the diffusion constant and ℓ is the characteristic diffusion length of the target.

The same argument applies when a product of gene k acts as a transcription factor. Let Ω_k be the effective volume in which that regulator is distributed, so its physical concentration is G_k/Ω_k. The corresponding diffusion-noise contribution is

$$\mathcal{N}_i^{(3)} = \sum_k \frac{\Omega_k}{D\ell}\left(R\frac{\partial f_i}{\partial G_k}\right)^2 G_k. \quad \text{(A.4)}$$

Recall the natural transcription-factor concentration scale

$$c_0 = \frac{N_{\max}}{D\ell\tau} \tag{A.5}$$

Using this, we define $c_{k,\max} = \frac{N_{\max}}{\Omega_k}$, the physical concentration corresponding to the normalized expression level $g_k = 1$. Combining the three independent source terms gives

$$\langle \Xi_i(t)\Xi_j(t')\rangle = \delta_{ij}\left[\frac{2G_i}{\tau} + \left(R\frac{\partial f_i}{\partial c}\right)^2 \frac{c}{D\ell} + \sum_k \frac{\Omega_k}{D\ell}\left(R\frac{\partial f_i}{\partial G_k}\right)^2 G_k\right]\delta(t-t'). \tag{A.6}$$

We now rescale to $g_i = G_i/N_{\max}$. Since $N_{\max} = R\tau$ and

$$\xi_i(t) = \frac{\tau}{N_{\max}}\Xi_i(t), \tag{A.7}$$

we obtain

$$\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\frac{\tau}{N_{\max}}\left[2\bar{g}_i + c_0 c\left(\frac{\partial f_i}{\partial c}\right)^2 + \sum_k \frac{c_0}{c_{k,\max}}\left(\frac{\partial f_i}{\partial g_k}\right)^2 \bar{g}_k\right]\delta(t-t'). \tag{A.8}$$

The factor c_0 in the external-input term follows from

$$\frac{1}{D\ell} = \frac{\tau}{N_{\max}}c_0, \tag{A.9}$$

while the internal-regulator factor follows from $\Omega_k/N_{\max} = 1/c_{k,\max}$.

For numerical work it is often convenient to use dimensionless concentrations

$$x = \frac{c}{c_0}, \qquad x_{k,\max} = \frac{c_{k,\max}}{c_0}. \tag{A.10}$$

Because $\partial f_i/\partial x = c_0\,\partial f_i/\partial c$, we have

$$c_0 c\left(\frac{\partial f_i}{\partial c}\right)^2 = x\left(\frac{\partial f_i}{\partial x}\right)^2, \tag{A.11}$$

and $c_0/c_{k,\max} = 1/x_{k,\max}$. Thus the equivalent dimensionless form is

$$N_{ij} = \delta_{ij}\frac{\tau}{N_{\max}}\left[2\bar{g}_i + x\left(\frac{\partial f_i}{\partial x}\right)^2 + \sum_k \frac{1}{x_{k,\max}}\left(\frac{\partial f_i}{\partial g_k}\right)^2 \bar{g}_k\right]. \tag{A.12}$$

The quantity $x_{k,\max}$ is the scale associated with an internally produced regulator and should not be confused with $x_{\max} = c_{\max}/c_0$, which specifies the accessible range of the external input.

B Derivation of information-optimizing input distributions

We follow the derivations of Tkačik et al. [2008]. Because $x = c/c_0$ is a one-to-one rescaling of the physical concentration, mutual information is invariant under this change of variables. We therefore perform the derivation using the physical input c.

For a fixed transcription-factor concentration, we assume that the output is narrowly distributed and approximately Gaussian. We use the notation $\mathcal{G}(g;\mu,\sigma^2)$ to denote a Gaussian with mean μ and variance σ^2, so we have

$$P(g \mid c) \approx \mathcal{G}\big(g;\bar{g}(c),\sigma_g^2(c)\big)\,. \tag{B.1}$$

Its conditional entropy is

$$H(g \mid c = c') = \log_2\left[\sqrt{2\pi e}\,\sigma_g(c')\right]. \tag{B.2}$$

In the small-noise limit, a locally monotonic mean response allows the change of variables

$$P(g) \approx \frac{P(c)}{|d\bar{g}/dc|}. \tag{B.3}$$

Hence

$$H(g) = -\int dc\,P(c)\log_2\left[\frac{P(c)}{|d\bar{g}/dc|}\right]. \tag{B.4}$$

Subtracting the average conditional entropy gives

$$\begin{aligned} I(c;g) &= -\int dc\,P(c)\log_2\left[\sqrt{2\pi e}\,P(c)\frac{\sigma_g(c)}{|d\bar{g}/dc|}\right] \\ &= -\int dc\,P(c)\log_2\left[\sqrt{2\pi e}\,P(c)\sigma_c(c)\right], \end{aligned} \tag{B.5}$$

where

$$\sigma_c(c) = \frac{\sigma_g(c)}{|d\bar{g}/dc|}. \tag{B.6}$$

We now maximize $I(c;g)$ subject to $\int_0^{c_{\max}} dc\, P(c) = 1$. Introducing a Lagrange multiplier λ, we get

$$\mathcal{L}[P] = -\int_0^{c_{\max}} dc\, P(c) \log_2\Big[\sqrt{2\pi e}\, P(c)\sigma_c(c)\Big] - \lambda\left[\int_0^{c_{\max}} dc\, P(c) - 1\right]. \tag{B.7}$$

The functional derivative is

$$\frac{\delta\mathcal{L}}{\delta P(c)} = -\log_2\Big[\sqrt{2\pi e}\,\sigma_c(c)P(c)\Big] - \frac{1}{\ln 2} - \lambda. \tag{B.8}$$

Setting this to zero shows that $P^*(c)\sigma_c(c)$ is constant, and therefore

$$P^*(c) = \frac{1}{Z}\frac{1}{\sigma_c(c)}, \qquad Z = \int_0^{c_{\max}} dc\, \frac{1}{\sigma_c(c)}. \tag{B.9}$$

Substitution into the mutual information gives

$$I^*(c;g) = \log_2\left(\frac{Z}{\sqrt{2\pi e}}\right). \tag{B.10}$$

The same expression is obtained if the derivation is carried out in $x = c/c_0$, because $dx = dc/c_0$ and $\sigma_x = \sigma_c/c_0$, so the factors of c_0 cancel in Z.

This argument also remains valid when a fluctuating non-cognate background is present. We write

$$c_{\rm nc}(t) = \bar{c}_{\rm nc} + \delta c_{\rm nc}(t), \tag{B.11}$$

where $\bar{c}_{\rm nc}$ is fixed during a capacity calculation. In the small-noise limit,

$$g(c, \bar{c}_{\rm nc} + \delta c_{\rm nc}) \approx \bar{g}(c;\bar{c}_{\rm nc}) + \left.\frac{\partial\bar{g}}{\partial c_{\rm nc}}\right|_{\bar{c}_{\rm nc}} \delta c_{\rm nc} + \xi(t), \tag{B.12}$$

where $\xi(t)$ collects the remaining Langevin noise sources. Under the Gaussian Langevin approximation, the linear combination of $\delta c_{\rm nc}$ and ξ is Gaussian, so after averaging over the non-cognate fluctuations,

$$P(g \mid c;\bar{c}_{\rm nc}) \approx \mathcal{G}\big(g;\bar{g}(c;\bar{c}_{\rm nc}), \sigma_g^2(c;\bar{c}_{\rm nc})\big). \tag{B.13}$$

The non-cognate factor therefore changes the mean response and total variance but does not introduce a second optimized input variable when $\bar{c}_{\rm nc}$ is fixed.

For the two-gene regulatory cascade, linearization around the steady state gives a linear system for $(\delta g_1, \delta g_2)$ driven by Gaussian Langevin forces. The joint distribution is therefore approximately multivariate Gaussian. Because a marginal of a multivariate Gaussian is Gaussian,

$$P(g_2 \mid c) \approx \mathcal{G}(g_2;\bar{g}_2(c), C_{22}(c)), \tag{B.14}$$

where $C_{22} = \sigma_{g_2}^2$ includes both the noise generated directly by the downstream gene and the noise propagated from the upstream gene. Thus, for the downstream channel, the same optimization gives $P^*(c) \propto 1/\sigma_c(c)$.

C Derivation of variance

C.1 Variance of simple one-gene binding

Writing $g_i = \bar{g}_i + \delta g_i$ and expanding f_i to the first order about the steady state gives

$$\tau\frac{d\delta g_i}{dt} \approx f_i(c;\{\bar{g}_j\}) + \sum_j \left.\frac{\partial f_i}{\partial g_j}\right|_{\bar{g}} \delta g - \bar{g}_i - \delta g_i + \xi_i(t). \tag{C.1}$$

Since the steady-state imposes $f_i(c;\{\bar{g}_j\}) = \bar{g}_i$, the above becomes

$$\tau\frac{d\delta g_i}{dt} \approx \sum_j \left.\frac{\partial f_i}{\partial g_j}\right|_{\bar{g}_j} \delta g_j - \delta g_i + \xi_i(t). \tag{C.2}$$

To simplify the equation, we work in the Fourier space by defining the Fourier transforms

$$\delta g_i(t) = \int \frac{d\omega}{2\pi}\, e^{-i\omega t}\delta\tilde{g}_i(\omega), \qquad \tilde{\xi}_i(t) = \int \frac{d\omega}{2\pi}\, e^{-i\omega t}\tilde{\xi}_i(\omega) \tag{C.3}$$

where $\delta\tilde{g}_i(\omega)$ and $\tilde{\xi}_i(\omega)$ denote the Fourier transforms of the gene-expression fluctuations and Langevin noise, respectively.

In the Fourier space, taking time derivatives corresponds to multiplication by $-i\omega$, so the dynamics become

$$(1 - i\omega\tau)\delta\tilde{g}_i - \sum_j \left.\frac{\partial f_i}{\partial g_j}\right|_{\bar{g}_j} \delta\tilde{g}_j = \tilde{\xi}_i. \tag{C.4}$$

For our simple one-gene one-transcription factor scenario, the generalized equation above simply becomes $(1 - i\omega\tau)\delta\tilde{g} = \tilde{\xi}$. To measure the contribution of fluctuations at frequency ω to the total variance in the output [Tkačik and Walczak, 2011], we introduce the power spectral density $S_g(\omega)$, defined through the relation

$$\langle \delta\tilde{g}(\omega)\delta\tilde{g}^*(\omega')\rangle = 2\pi\delta(\omega - \omega')S_g(\omega), \tag{C.5}$$

where $\delta\tilde{g}^*$ is the complex conjugate of $\delta\tilde{g}$. From Tkačik and Walczak [2011] and Tkačik et al. [2008], we can calculate $\sigma_g(c)$ through the equation

$$\sigma_g(c)^2 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_g(\omega). \tag{C.6}$$

By substituting $\delta\tilde{g} = \tilde{\xi}/(1 - i\omega\tau)$ and $\delta\tilde{g}^* = \tilde{\xi}^*/(1 + i\omega'\tau)$, we find

$$\begin{aligned} \langle \delta\tilde{g}(\omega)\delta\tilde{g}^*(\omega')\rangle &= \frac{\langle\tilde{\xi}(\omega)\tilde{\xi}^*(\omega')\rangle}{(1 + i\omega'\tau)(1 - i\omega\tau)} \\ &= \frac{2\pi\delta(\omega - \omega')N}{1 + \omega^2\tau^2} \\ S_g(\omega) &= \frac{N}{1 + \omega^2\tau^2}. \end{aligned} \tag{C.7}$$

From (C.1), we get

$$\sigma_g^2(c) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{N}{1 + \omega^2\tau^2} = \frac{N}{2\tau}. \tag{C.8}$$

C.2 Variance of the regulatory cascade motif

For multi-gene systems such as this one, it is useful to write the linearized dynamics in matrix form in Fourier space. We define

$$\phi_{ij} = \left.\frac{\partial f_i}{\partial g_j}\right|_{\bar{\mathbf{g}}}. \tag{C.9}$$

Then, (C.1) becomes

$$(1 - i\omega\tau)\delta\tilde{g}_i - \sum_j \phi_{ij}\delta\tilde{g}_j = \tilde{\xi}_i. \tag{C.10}$$

For a regulatory network of multiple genes, we may gather all the equations in the Fourier space into a matrix system where

$$A(\omega)\,\delta\tilde{\mathbf{g}}(\omega) = \tilde{\boldsymbol{\xi}}(\omega),$$

$$A(\omega) = \begin{pmatrix} 1 - i\omega\tau - \phi_{11} & -\phi_{12} & \cdots \\ -\phi_{21} & 1 - i\omega\tau - \phi_{22} & \cdots \\ \vdots & \vdots & \ddots, \end{pmatrix}$$

$$\delta\tilde{\mathbf{g}}(\omega) = \begin{pmatrix} \delta\tilde{g}_1(\omega) \\ \delta\tilde{g}_2(\omega) \\ \vdots \end{pmatrix}, \qquad \tilde{\boldsymbol{\xi}}(\omega) = \begin{pmatrix} \tilde{\xi}_1(\omega) \\ \tilde{\xi}_2(\omega) \\ \vdots \end{pmatrix} \tag{C.11}$$

In our two-gene case, we find

$$A = \begin{pmatrix} 1 - i\omega\tau & 0 \\ -\phi_{21} & 1 - i\omega\tau \end{pmatrix} \tag{C.12}$$

as the terms $\phi_{11}, \phi_{12}, \phi_{22}$ are simply 0.

To derive the variance directly, we use the equation

$$\frac{1}{\sigma_c^2(c)} = \sum_{ij} \frac{d\bar{g}_i}{dc} C_{ij}^{-1} \frac{d\bar{g}_j}{dc}, \tag{C.13}$$

where C_{ij} is an element of the covariance matrix [Tkačik and Walczak, 2011]. Tkačik and Walczak [2011] also give the relation

$$C = \int \frac{d\omega}{2\pi} A^{-1}(\omega) N [A^{-1}(\omega)]^\dagger, \tag{C.14}$$

where $A^\dagger$ denotes the Hermitian conjugate of A and N is the noise spectral density matrix. From (C.2), we have $\det A = (1 - i\omega\tau)^2$. Thus,

$$\begin{aligned} A^{-1}(\omega) &= \frac{1}{(1 - i\omega\tau)^2} \begin{pmatrix} 1 - i\omega\tau & 0 \\ \phi_{21} & 1 - i\omega\tau \end{pmatrix} \\ [A^{-1}(\omega)]^\dagger &= \frac{1}{(1 + i\omega\tau)^2} \begin{pmatrix} 1 + i\omega\tau & \phi_{21} \\ 0 & 1 + i\omega\tau \end{pmatrix}. \end{aligned} \tag{C.15}$$

We also assume that there is no cross noise spectral density, so that

$$N = \begin{pmatrix} N_1 & 0 \\ 0 & N_2 \end{pmatrix}. \tag{C.16}$$

Multiplying the matrices out, we have

$$A^{-1}(\omega)N[A^{-1}(\omega)]^\dagger = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

$$S_{11}(\omega) = \frac{N_1}{1+\omega^2\tau^2},$$

$$S_{22}(\omega) = \frac{N_2}{1+\omega^2\tau^2} + \frac{N_1\phi_{21}^2}{(1+\omega^2\tau^2)^2}, \tag{C.17}$$

$$S_{12}(\omega) = \frac{N_1\phi_{21}(1-i\omega\tau)}{(1+\omega^2\tau^2)^2},$$

$$S_{21}(\omega) = \frac{N_1\phi_{21}(1+i\omega\tau)}{(1+\omega^2\tau^2)^2}.$$

Integrating over frequency, we find the covariance matrix

$$\begin{aligned} C &= \int \frac{d\omega}{2\pi} A^{-1}(\omega)N[A^{-1}(\omega)]^\dagger \\ &= \begin{pmatrix} \frac{N_1}{2\tau} & \frac{N_1\phi_{21}}{4\tau} \\ \frac{N_1\phi_{21}}{4\tau} & \frac{N_2}{2\tau} + \frac{N_1\phi_{21}^2}{4\tau} \end{pmatrix}. \end{aligned} \tag{C.18}$$

When considering g_2 as the sole output of the regulatory system, the multivariable covariance matrix reduces to the scalar variance $C_{22} = \sigma_{g_2}(c)^2$. Thus, (C.2) becomes

$$\begin{aligned} \sigma_c(c)^2 &= \frac{C_{22}}{\left(\frac{d\bar{g}_2}{dc}\right)^2} \\ &= \frac{2N_2 + N_1\phi_{21}^2}{4\tau\left(\frac{d\bar{g}_2}{dc}\right)^2}. \end{aligned} \tag{C.19}$$

We can then use the steady state assumption $\bar{g}_2(c) = f_2(g_1)$ to calculate ϕ_{21} and $d\bar{g}_2/dc$.

D Non-cognate transcription factor MWC model

We next extend the MWC framework to describe competition between cognate and non-cognate transcription factors. As in the standard MWC model, the regulatory region can exist in either an "on" or an "off" conformation. For a single site in the "on" state, the statistical weights of the empty, cognate-bound, and non-cognate-bound configurations are respectively

$$1, \qquad \frac{c}{K_{d,\mathrm{on}}}, \qquad \frac{c_{\mathrm{nc}}}{K_{\mathrm{nc,on}}}.$$

The single-site partition factor in the "on" state is therefore

$$1 + \frac{c}{K_{d,\mathrm{on}}} + \frac{c_{\mathrm{nc}}}{K_{\mathrm{nc,on}}}. \tag{D.1}$$

Assuming that the n sites are identical and independent conditional on the global promoter conformation, the total statistical weight of all configurations in the "on" state is

$$Z_{\mathrm{on}} = \left(1 + \frac{c}{K_{d,\mathrm{on}}} + \frac{c_{\mathrm{nc}}}{K_{\mathrm{nc,on}}}\right)^n. \tag{D.2}$$

Similarly, the statistical weight of the OFF state is

$$Z_{\mathrm{off}} = e^{\Delta\varepsilon}\left(1 + \frac{c}{K_{d,\mathrm{off}}} + \frac{c_{\mathrm{nc}}}{K_{\mathrm{nc,off}}}\right)^n, \tag{D.3}$$

where $\Delta\varepsilon$ represents the energetic bias toward the "off" conformation.

We now make two simplifying assumptions. First, we assume the limit $c/K_{\mathrm{d,\ off}} \ll 1$. Furthermore, $K_{d,\mathrm{on}} = K_d$ is simply the dissociation constant. Next, we assume that the "on" and "off" conformations of the promoter do not effect non-cognate binding, i.e. $K_{\mathrm{nc,on}} = K_{\mathrm{nc,off}} = K_{\mathrm{nc}}$. Under these assumptions, the two partition functions become

$$Z_{\mathrm{on}} = \left(1 + \frac{c}{K_d} + \frac{c_{\mathrm{nc}}}{K_{\mathrm{nc}}}\right)^n, \tag{D.4}$$

$$Z_{\mathrm{off}} = e^{\Delta\varepsilon}\left(1 + \frac{c_{\mathrm{nc}}}{K_{\mathrm{nc}}}\right)^n. \tag{D.5}$$

If transcriptional activity is proportional to the probability that the regulatory region is in the "on" state, the input–output function becomes

$$f(c, c_{\mathrm{nc}}) = \frac{Z_{\mathrm{on}}}{Z_{\mathrm{on}} + Z_{\mathrm{off}}}, \tag{D.6}$$

which gives

$$f(c, c_{\rm nc}) = \frac{\left(1 + \dfrac{c}{K_d} + \dfrac{c_{\rm nc}}{K_{\rm nc}}\right)^n}{\left(1 + \dfrac{c}{K_d} + \dfrac{c_{\rm nc}}{K_{\rm nc}}\right)^n + e^{\Delta\varepsilon}\left(1 + \dfrac{c_{\rm nc}}{K_{\rm nc}}\right)^n}. \tag{D.7}$$

As in the single-transcription-factor MWC model, this expression can be written in the logistic form

$$f(c, c_{\rm nc}) = \frac{1}{1 + e^{F(c, c_{\rm nc})}}, \tag{D.8}$$

where

$$\begin{aligned} F(c, c_{\rm nc}) = \Delta\varepsilon &+ n \ln\left(1 + \frac{c_{\rm nc}}{K_{\rm nc}}\right) \\ &- n \ln\left(1 + \frac{c}{K_d} + \frac{c_{\rm nc}}{K_{\rm nc}}\right). \end{aligned} \tag{D.9}$$

References

Uri Alon. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, 8(6):450–461, June 2007.

Justin Crocker, Nobuaki Abe, Luca Rinaldi, Alistair P. McGregor, Nicholas Frankel, Sheng Wang, Atheer Alsawadi, Paolo Valenti, Sandra Plaza, François Payre, Richard S. Mann, and David L. Stern. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*, 160(1–2):191–203, 2015. ISSN 0092-8674. doi: 10.1016/j.cell.2014.11.041. URL https://www.cell.com/cell/fulltext/S0092-8674(14)01518-9.

Justin Crocker, Ella Preger-Ben Noon, and David L. Stern. Chapter twenty-seven - the soft touch: Low-affinity transcription factor binding sites in development and evolution. In Paul M. Wassarman, editor, *Essays on Developmental Biology, Part B*, volume 117 of *Current Topics in Developmental Biology*, pages 455–469. Academic Press, 2016. doi: https://doi.org/10.1016/bs.ctdb.2015.11.018. URL https://www.sciencedirect.com/science/article/pii/S0070215315001842.

Valer Gotea, Axel Visel, John M Westlund, Marcelo A Nobrega, Len A Pennacchio, and Ivan Ovcharenko. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.*, 20(5):565–577, May 2010.

Archibald Vivian Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *The Journal of Physiology*, 40: iv–vii, 1910. doi: 10.1113/jphysiol.1910.sp001386.

Ying Kao-Huang, Arnold Revzin, Andrew P. Butler, Pamela O'Conner, Daniel W. Noble, and Peter H. Von Hippel. Nonspecific dna binding of genome-regulating proteins as a biological control mechanism: Measurement of dna-bound escherichia coli lac repressor in vivo. *Proceedings of the National Academy of Sciences*, 74(10):4228–4232, 1977. doi: 10.1073/pnas.74.10.4228. URL https://www.pnas.org/doi/abs/10.1073/pnas.74.10.4228.

Lan Ma, John Wagner, John Jeremy Rice, Wenwei Hu, Arnold J. Levine, and Gustavo A. Stolovitzky. A plausible model for the digital response of p53 to dna damage. *Proceedings of the National Academy of Sciences*, 102(40):14266–14271, 2005. doi: 10.1073/pnas.0501352102. URL https://www.pnas.org/doi/abs/10.1073/pnas.0501352102.

Sara D. Mahdavi, Gabriel L. Salmon, Patill Daghlian, Hernan G. Garcia, and Rob Phillips. Flexibility and sensitivity in gene regulation out of equilibrium. *Proceedings of the National Academy of Sciences*, 121(46):e2411395121, 2024. doi: 10.1073/pnas.2411395121. URL https://www.pnas.org/doi/abs/10.1073/pnas.2411395121.

J Monod, J Wyman, and J P Changeux. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.*, 12(1):88–118, May 1965.

Julian Naderi, Alexandre P. Magalhães, Gözde Kibar, Grégoire Stik, Yaotian Zhang, Sebastian D. Mackowiak, Hannah M. Wieler, Francesca Rossi, René Buschow, Marie Christou-Kent, Marc Alcoverro-Bertran, Thomas Graf, Martin Vingron, and Denes Hnisz. An activity-specificity trade-off encoded in human transcription factors. *Nature Cell Biology*, 26(8):1309–1321, July 2024. doi: 10.1038/s41556-024-01411-0. URL https://www.nature.com/articles/s41556-024-01411-0.pdf.

Béla Novák and John J Tyson. Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.*, 9 (12):981–991, December 2008.

Rob Phillips and Nigel Orme. *The Molecular Switch: Signaling and Allostery*. Princeton University Press, 2020. ISBN 9780691200248. URL http://www.jstor.org/stable/j.ctvx5w8pf.

Amir Shahein, Maria López-Malo, Ivan Istomin, Evan J Olson, Shiyu Cheng, and Sebastian J Maerkl. Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance. *Nat. Commun.*, 13 (1):5273, September 2022.

Gasper Tkacik, Curtis G Callan, Jr, and William Bialek. Information capacity of genetic regulatory elements. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 78(1 Pt 1):011910, July 2008a.

Gasper Tkacik, Thomas Gregor, and William Bialek. The role of input noise in transcriptional regulation. *PLoS One*, 3(7):e2774, July 2008b.

Ga šper Tkačik, Curtis G. Callan, and William Bialek. Information capacity of genetic regulatory elements. *Phys. Rev. E*, 78:011910, Jul 2008. doi: 10.1103/PhysRevE.78.011910. URL https://link.aps.org/doi/10.1103/PhysRevE.78.011910.

Ga šper Tkačik, Aleksandra M. Walczak, and William Bialek. Optimizing information flow in small genetic networks. *Phys. Rev. E*, 80:031920, Sep 2009. doi: 10.1103/PhysRevE.80.031920. URL https://link.aps.org/doi/10.1103/PhysRevE.80.031920.

Gašper Tkačik and Aleksandra M Walczak. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter*, 23(15):153102, apr 2011. doi: 10.1088/0953-8984/23/15/153102. URL https://doi.org/10.1088/0953-8984/23/15/153102.

Gašper Tkačik, Curtis G. Callan, and William Bialek. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences*, 105(34):12265–12270, 2008. doi: 10.1073/pnas.0806077105. URL https://www.pnas.org/doi/abs/10.1073/pnas.0806077105.

Aleksandra M. Walczak, Ga šper Tkačik, and William Bialek. Optimizing information flow in small genetic networks. ii. feed-forward interactions. *Phys. Rev. E*, 81:041905, Apr 2010. doi: 10.1103/PhysRevE.81.041905. URL https://link.aps.org/doi/10.1103/PhysRevE.81.041905.

Zeba Wunderlich and Leonid A Mirny. Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet*, 25(10):434–440, October 2009.

Nicolae Radu Zabet and Boris Adryan. The effects of transcription factor competition on gene regulation. *Frontiers in Genetics*, Volume 4 - 2013, 2013. ISSN 1664-8021. doi: 10.3389/fgene.2013.00197. URL https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2013.00197.

Etay Ziv, Ilya Nemenman, and Chris H Wiggins. Optimal signal processing in small stochastic biochemical networks. *PLoS One*, 2(10):e1077, October 2007.