



## Optimizing information flow in small genetic networks. IV. Spatial coupling

Thomas R. Sokolowski\* and Gašper Tkačik†

*Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria*

(Received 16 January 2015; published 15 June 2015)

We typically think of cells as responding to external signals independently by regulating their gene expression levels, yet they often locally exchange information and coordinate. Can such spatial coupling be of benefit for conveying signals subject to gene regulatory noise? Here we extend our information-theoretic framework for gene regulation to spatially extended systems. As an example, we consider a lattice of nuclei responding to a concentration field of a transcriptional regulator (the input) by expressing a single diffusible target gene. When input concentrations are low, diffusive coupling markedly improves information transmission; optimal gene activation functions also systematically change. A qualitatively different regulatory strategy emerges where individual cells respond to the input in a nearly steplike fashion that is subsequently averaged out by strong diffusion. While motivated by early patterning events in the *Drosophila* embryo, our framework is generically applicable to spatially coupled stochastic gene expression models.

DOI: [10.1103/PhysRevE.91.062710](https://doi.org/10.1103/PhysRevE.91.062710)

PACS number(s): 87.16.Yc, 87.16.Xa, 87.18.Tt, 87.18.Gh

### I. INTRODUCTION

Signals encoded by spatial or temporal concentration profiles of certain proteins play a crucial role in communicating information within and between cells. Such signals are established and read out by a vast variety of gene regulatory networks. In this process, however, information flow is limited by the randomness associated with gene regulatory interactions—essentially chemical reactions between different species of signaling molecules and DNA—taking place at very low copy numbers [1]. While we increasingly understand distinct strategies of noise control in biological systems [2–6], it remains largely unclear how nature orchestrates these strategies to maximize information flow.

A recent proposal takes the idea of information flow in biological networks seriously, as formalized by Shannon's information theory, and hypothesizes that at least some of the structure of genetic regulatory networks could be derived mathematically, by maximizing information transmission that such networks can sustain subject to biophysically realistic constraints [7]. The feasibility of information as the metric of network performance was recently established by demonstrating that morphogen patterns in early fly development transmit information with an accuracy close to the physical limits [8,9]. Theoretically, this optimization principle was applied to study information flow in different paradigmatic gene-regulatory scenarios, including single and multiple target genes regulated by a single input [10–12], feed-forward cross regulation [12], autoactivation [13], multistate promotor switching [14], and time-dependent inputs [15–19]. These attempts yielded important insights into optimal strategies with which single cells can reliably respond to external stimuli. The issue that to date has not received attention, however, is that many cells or nuclei can respond to a spatially distributed signal collectively, by exchanging information among themselves, e.g., via diffusion of signaling proteins. Prominent examples are the specification of emerging body tissues during embryo

development [20–25], the aggregation response in colonies of amebas [26,27], and collective chemosensing in cell colonies and tissues [28,29]. Importantly, spatial aspects not only add another layer of complexity to the distributions of input and output signals of gene-regulatory networks, but they can also markedly alter noise levels, and thus information capacity, through spatial averaging [30–35]. A truly predictive theoretical framework therefore must take into account the spatial setting, where multiple regulatory networks can read out the inputs at various locations and exchange information between themselves locally.

Here we extend the existing theory of optimal gene regulatory networks to such a spatial setting. We construct a spatial-stochastic model of gene regulation that takes into account diffusive coupling between neighboring reaction volumes and the relevant sources of noise. We derive simple expressions for the stationary means and variances of input-activated local protein levels as a function of their regulatory parameters and diffusion rate. From this we compute information transmission in various spatial scenarios and systematically explore how transmission can be optimized by changing key system parameters. As a motivating example we study a model variant representative of the bicoid-hunchback system in early fruit fly embryogenesis, which enables differentiating cells to acquire position-dependent fates with high reliability and reproducibility in spite of stochasticity in the underlying biophysical processes [32,36–42]. Our approach thus specifically optimizes positional information, a notion that has been extensively discussed throughout developmental biology [43–47] but only recently rigorously formalized in mathematical terms using information theory [48]; the proposed formalism, however, generalizes to other spatially coupled gene regulatory setups.

We find that spatial coupling can markedly improve information transmission when the dominant source of noise in gene regulation is on the input side, e.g., due to the random arrival of the regulatory molecules (transcription factors) to their binding sites on the DNA. This is in accordance with the known fact that spatial averaging can remove super-Poissonian contributions to the noise [30,31]. Moreover, with diffusion, the maximal information capacity reached upon optimization becomes increasingly invariant to the spatial details of the input

\*thomas.sokolowski@ist.ac.at

†gasper.tkacik@ist.ac.at

signal. Finally, in a defined noise regime that we detail below, we observe the emergence of a qualitatively different optimal regulatory strategy where diffusion generates graded, and thus informative, spatial profiles of regulated genes despite their sharp thresholdlike activation by the regulatory inputs.

## II. METHODS

Here we construct a model in which we can analytically calculate the amount of information that expression levels of diffusible proteins locally encode about the position itself, if these proteins are expressed in response to a concentration field of regulatory input molecules. The outline of our approach is as follows. We (i) define a generic spatial-stochastic model of gene regulation that explicitly accounts for diffusive signaling between adjacent volumes that collectively interpret a spatially distributed input signal (Sec. II A). Starting from a set of coupled Langevin equations, in Sec. II B we (ii) derive general analytic expressions for the stationary means and variances of the regulated protein levels and provide intuitive insight for the role of diffusion prior to specifying any gene regulatory details. We then (iii) impose realistic expressions for the gene regulatory function and noise, based on well-established biochemical models (Sec. II C). Section II D illustrates how we (iv) quantify the encoding of positional information, using information theory and the stationary moments derived in (ii). Finally, we vary the parameters of the model to find the optimal regulatory strategies in the spatially coupled setup, which we present in Sec. III.

### A. Spatial-stochastic model of gene regulation

Figure 1 shows a schematic of the spatially coupled gene regulation model that we study throughout this paper. We consider a two-dimensional lattice of  $N_x \times N_y$  reaction volumes with equal spacing  $\delta$  between the volumes. In this work we particularly focus on a cylindrical geometry of the lattice with periodic boundary conditions along the circumference ( $y$  direction) and reflective boundary conditions in the axial ( $x$ ) direction. While this geometry deliberately mimics the syncytial state of the developing embryo of the fruit fly *Drosophila*, our model could be easily adapted to any arbitrary discrete arrangement of equally spaced reaction volumes. We will initially consider the special case  $N_y = 1$  as the one-dimensional (1D) model, while referring to the case  $N_y > 1$  as the 2D model.

Each volume contains the promotor of an identical gene controlled by a single transcription factor, whose concentration we denote by  $c$ . The copy number of proteins expressed from this gene of interest will be denoted by  $G$ , while  $g = G/N_{\max}$  will refer to the copy number normalized by the maximal average expression level  $N_{\max}$  at full induction. At the core of our model is the regulatory function  $f(c)$ , which determines how the activator concentration  $c$  maps to the effective synthesis rate of  $G$  proteins. We will also refer to  $c$  as the input and to  $G$  and  $g$  as the output. The  $G$  molecules are degraded with a constant rate  $1/\tau$ , where  $\tau$  is the average protein lifetime. For simplicity we treat transcription and translation as a single production step with maximal production rate  $r_{\max}$  reached at full activation. The local total production rate  $r_i \equiv$

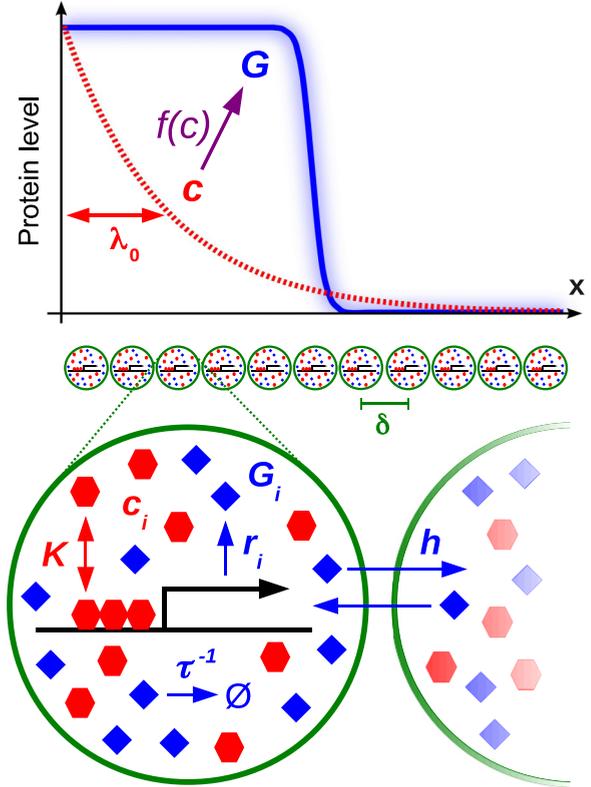


FIG. 1. (Color online) Model schematic. Our spatial-stochastic model of gene regulation consists of a discrete set of identical reaction volumes (indicated by  $i$ ) arranged in a regular lattice with lattice spacing  $\delta$ . While here for clarity we only draw a 1D lattice, we also study a 2D model with periodic boundary conditions and our framework supports arbitrary coordination numbers. The volumes collectively sense a spatial signal  $c(x)$  that maps the position  $x$  to an input concentration  $c$  (red), creating a spatial distribution of output proteins  $G$  (blue). While our model allows for arbitrary input functions  $c(x)$ , in this work we focus on exponential input gradients with length constant  $\lambda\gamma_0$ , where  $\gamma_0$  is a typical length scale and  $\lambda$  a dimensionless parameter (cf. Sec. II D). Each volume contains an identical promotor activated with probability  $f(c_i)$ , where  $f(c)$  is the regulatory function and  $c_i$  the activator concentration in volume  $i$ . This results in a local production rate  $r_i = r(c_i)$  and local levels  $G_i$  of the product protein;  $G$  proteins are degraded with a rate  $1/\tau$  and can hop to the neighboring volumes with a rate  $h = D/\delta^2$ , where  $D$  is their diffusion coefficient. The regulatory function  $f(c)$  is characterized by the activation threshold  $K$  and the regulatory cooperativity  $H$ , as detailed in Sec. II C.

$r(c(\mathbf{x}_i))$  is composed of a basal expression rate  $r_b$  capturing leaky production of the gene and a part proportional to the activation level  $r_i = r_b + (r_{\max} - r_b)f(c(\mathbf{x}_i))$ . Throughout this work we use a very small basal rate of expression  $r_b = 0.01/\tau$ , as explained in Appendix D. Importantly, each volume can exchange the product molecules  $G$  with its neighboring volumes via diffusion. Here we approximate this exchange with an effective hopping rate  $h = D/\delta^2$ , where  $D$  is the diffusion coefficient of  $G$  molecules.

Denoting the position of volume  $i$  by  $\mathbf{x}_i$ , we can write down a stochastic equation for the combined production-

degradation-diffusion process in volume  $i$  as follows:

$$\begin{aligned} \partial_t G_i &= r(c(\mathbf{x}_i)) - \frac{1}{\tau} G_i \\ &\quad - \frac{D}{\delta^2} \sum_{n \in N(i)} (G_i - G_n) + \sum_k \Gamma_{ik} \eta_k. \end{aligned} \quad (1)$$

Here  $G_i$  denotes the number of  $G$  proteins in the volume, while the first and second terms in the equation describe their production and degradation, respectively. The third term is a discretized Laplacian accounting for the diffusive exchange of  $G$  molecules between volume  $i$  and its nearest-neighbor volumes  $n \in N(i)$ . The function  $c(\mathbf{x})$  maps positions  $\mathbf{x}_i$  to local transcription factor concentrations  $c_i$ . In this work we consider the case in which  $c(\mathbf{x})$  is invariant in the  $y$  direction, i.e.,  $c_i \equiv c(\mathbf{x}_i) = c(x(i))$ ; we specify the functional form of  $c(x)$  at the end of this section. The noise term  $\sum_k \Gamma_{ik} \eta_k$  comprises all noise sources  $\eta_k$  that act on the protein level  $G_i$ , including those outside of volume  $i$ . We assume that all fluctuations are well represented by nonmultiplicative zero-mean Gaussian white noise and compute the relevant noise powers  $\Gamma_{ik}$  below.

### B. Means and variances with spatial coupling

The set of coupled Langevin equations [Eq. (1)] is a special case of what is known as the  $N$ -unit generalized Langevin model [49,50] in other contexts. To compute the steady-state means and variances we employ a technique based on Itô calculus that allows us to convert the set of coupled stochastic equations for the variables  $G_i$  into a set of ordinary differential equations for their moments that hold exactly. This can be performed for an arbitrary discrete coupling network and even for the case of multiplicative noise (not considered here) [49–51].

We can write down the equations of motion for the mean of the normalized local output  $g_i = G_i/N_{\max} = G_i/r_{\max}\tau$  and the (co)variances of its fluctuations  $\delta g_i \equiv g_i - \langle g_i \rangle$  as follows [49–51]:

$$\partial_t \langle g_i \rangle = \langle F(g_i, c_i) \rangle - h \sum_{n_i \in N(i)} \langle g_i - g_{n_i} \rangle, \quad (2)$$

$$\begin{aligned} \partial_t \langle \delta g_i \delta g_j \rangle &= \langle \delta g_i F(g_j, c_j) \rangle + \langle \delta g_j F(g_i, c_i) \rangle \\ &\quad + h \left\langle \delta g_i \sum_{n_j \in N(j)} (g_{n_j} - g_j) \right\rangle \\ &\quad + h \left\langle \delta g_j \sum_{n_i \in N(i)} (g_{n_i} - g_i) \right\rangle \\ &\quad + \sum_k \langle \Gamma_{ik} \Gamma_{jk} \rangle. \end{aligned} \quad (3)$$

Here the function  $F(g_i, c_i) = \frac{1}{\tau} (\frac{r_i}{r_{\max}} - g_i)$  groups the production and degradation terms and depends on the regulatory function  $f(c)$ . The  $n_i$  and  $n_j$  sums run over the nearest-neighbor volumes of  $i$  and  $j$ , respectively. Note that in the  $k$  sum running over all noise sources in the system usually most terms will be zero.

The hitherto arbitrary noise powers  $\Gamma_{ik}$  can be written out more explicitly by listing all noise sources that affect volume

$i$  in the following way:

$$\begin{aligned} \sum_k \Gamma_{ik} \eta_k &= \mathcal{G}_i(c_i, \mathbf{g}) \gamma_i - \sum_{n \in N(i)} \mathcal{D}_{(i \rightarrow n)} \xi_{(i \rightarrow n)} \\ &\quad + \sum_{n \in N(i)} \mathcal{D}_{(n \rightarrow i)} \xi_{(n \rightarrow i)}. \end{aligned} \quad (4)$$

Herein  $\gamma_i$  is the noise in the combined process of random regulation, production, and degradation and  $\xi_{(i \rightarrow j)}$  is the noise caused by random hopping from volume  $i$  to  $j$ . A crucial step in specifying the expression above is to write the noise powers of the incoming and outgoing hopping processes with different signs, because each hopping event causing a negative fluctuation in the volume of origin concomitantly causes a positive fluctuation in the volume of arrival. The hopping thus is a birth process when seen from the volume of arrival, and a death process when seen from the volume of origin. This intuitive argument can be derived more rigorously from the Langevin equation following the method portrayed by Gillespie [52] and is quantitatively confirmed by stochastic simulations, employing, e.g., the next-subvolume method [53,54].

Using Eq. (4) we can rewrite the noise term  $\sum_k \langle \Gamma_{ik} \Gamma_{jk} \rangle$  in Eq. (3) for the variances (case  $i = j$ ) and nearest-neighbor covariances [case  $i \in N(j)$ ] as follows:

$$\begin{aligned} \sum_k \langle \Gamma_{ik} \Gamma_{jk} \rangle &= \langle \mathcal{G}_i^2(c, \mathbf{g}) \rangle + \sum_{n \in N(i)} (\mathcal{D}_{(i \rightarrow n)}^2 + \mathcal{D}_{(n \rightarrow i)}^2) \\ &\equiv \mathcal{N}_{ii}^2 \quad \text{for } i = j, \end{aligned} \quad (5)$$

$$\begin{aligned} \sum_k \langle \Gamma_{ik} \Gamma_{jk} \rangle &= \underbrace{\langle \Gamma_{ik} \Gamma_{jk} \rangle}_{\text{hopping } (i \rightarrow j)} + \underbrace{\langle \Gamma_{ik'} \Gamma_{jk'} \rangle}_{\text{hopping } (j \rightarrow i)} \\ &= -\langle \mathcal{D}_{(i \rightarrow j)}^2 \rangle - \langle \mathcal{D}_{(j \rightarrow i)}^2 \rangle \\ &\equiv \mathcal{N}_{ij}^2 \quad \text{for } i \in N(j). \end{aligned} \quad (6)$$

To obtain the stationary means and covariances of the protein number  $g_i$ , we solved Eqs. (2) and (3) in the steady state by setting their left-hand sides to zero ( $\partial_t(\dots) = 0$ ). After substituting  $g_i = \langle g_i \rangle + \delta g_i$ , we regroup covariances  $\langle \delta g_i \delta g_j \rangle$  and means  $\langle g_i \rangle$  and relate them to the respective quantities in the neighboring volumes (see Appendix A). In this context, we introduce the pragmatic assumption that only concentrations in volumes that are nearest neighbors on the lattice have significant correlations, i.e., we enforce

$$\langle \delta g_i \delta g_j \rangle = 0 \quad \text{for } j \notin N(i). \quad (7)$$

While this short-correlation assumption markedly facilitates both analytical progress and computation speed and yields formulas that are intuitive to interpret, our results only change marginally when the assumption is released (see Sec. III D). For brevity, in the following we write

$$\bar{g}_i \equiv \langle g_i \rangle, \quad \sigma_i^2 \equiv \langle \delta g_i^2 \rangle, \quad C_{ij} \equiv \langle \delta g_i \delta g_j \rangle \quad (8)$$

for the steady-state means, variances, and covariances, respectively. Treating the cases  $i = j$  and  $i \in N(j)$  in Eq. (3) separately, we arrive at the following set of coupled equations

for these quantities (cf. Appendix A):

$$\begin{aligned}\bar{g}_i &= \frac{\tau}{1 + 2d\Delta} \left( \hat{r}_i + h \sum_{n_i \in N(i)} \bar{g}_{n_i} \right) \\ &\equiv T\hat{r}_i + \Lambda^2 \sum_{n_i \in N(i)} \bar{g}_{n_i},\end{aligned}\quad (9)$$

$$\sigma_i^2 = \Lambda^2 \sum_{n_i \in N(i)} C_{in_i} + \frac{T}{2} \mathcal{N}_{ii}^2, \quad (10)$$

$$C_{ij} = \frac{\Lambda^2}{2} (\sigma_i^2 + \sigma_j^2) + \frac{T}{2} \mathcal{N}_{ij}^2. \quad (11)$$

Here we abbreviate the normalized production rate  $\hat{r}_i = [\bar{g}_b + (1 - \bar{g}_b)f(c_i)]/\tau$ , where  $\bar{g}_b \equiv r_b/r_{\max}$  equals the basal expression level in normalized units, and  $d$  is the lattice dimension, or half of the coordination number of a reaction volume. We introduce the dimensionless diffusion constant  $\Delta \equiv D/D_0$ , which is equal to the diffusion constant measured in the natural unit  $D_0 = \delta^2/\tau$ ; note that  $\Delta = h\tau$ . We will also refer to  $\Delta$  as the spatial coupling.

The above equations define two effective parameters: (i) an effective residence time  $T = \tau/(1 + 2d\Delta)$ , reflecting the fact that with diffusion proteins are taken out from the reaction volume faster than through regular degradation in the absence of diffusion, and (ii) the (dimensionless) mixing parameter  $\Lambda = \sqrt[4]{hT} = \sqrt[4]{DT}/\delta = \sqrt[4]{\Delta/(1 + 2d\Delta)}$ , which equals the distance traveled via diffusion during the time  $T$ , measured in units of the lattice spacing  $\delta$ . These quantities capture the essential effect of diffusive coupling between the stochastic processes in neighboring reaction volumes: With increasing coupling ( $\Delta > 0 \Rightarrow \Lambda^2 > 0$ ), the mean output level  $\bar{g}_i$  is increasingly set by the mean levels in the neighboring volumes  $\bar{g}_{n_i}$ , whereas the contribution of the local production to the apparent copy number in the volume is reduced ( $T < \tau$ ). An analogous interpretation holds for  $\sigma_i^2$  if  $\frac{1}{2}\mathcal{N}_{ii}^2$  is interpreted as a noise production term.

We verified that Eqs. (9)–(11) correctly reproduce the known result that all super-Poissonian noise is attenuated in the Poissonian limit, meaning that (in non-normalized units) the variance becomes equal to the mean for infinitely strong coupling  $\Delta \rightarrow \infty$  [30,31]. This holds in both one and two dimensions and with or without the short-correlation assumption.

Equations (9)–(11) can be solved for the variables  $\bar{g}_i$ ,  $\sigma_i^2$ , and  $C_{ij}$  after specifying the detailed forms of the regulatory function and noise strength and imposing meaningful boundary conditions. If the regulatory function and noise do not contain terms nonlinear in the variables of interest, Eqs. (9)–(11) form a set of coupled linear equations that can be solved by simple matrix inversion. Also notice that we may use Eq. (11) to simplify Eq. (10) further if we are only interested in solving for the variances  $\sigma_i^2$  (as in this work); we show the respective formulas for the 1D and 2D models in Appendix B.

### C. Regulatory functions and noise powers

To describe the probability of the promotor being in its active state given that the concentration of the transcription

factor is  $c$ , we choose a simple Hill-type regulation model

$$f(c) = \frac{c^H}{c^H + K^H}, \quad (12)$$

where  $H$  is the Hill coefficient, quantifying the cooperativity of the activation process, and  $K$  is the activation threshold, i.e., the concentration at which half activation occurs:  $f(K) = 1/2$ .

To complete our analytical model we need to accurately specify the noise powers  $\langle \mathcal{G}_i^2 \rangle$  and  $\langle \mathcal{D}_{(i \rightarrow j)}^2 \rangle$  appearing in the noise terms  $\mathcal{N}_{ii}^2$  and  $\mathcal{N}_{ij}^2$ . In the steady state, all of these noise powers only depend on the mean levels of the transcription factor  $\{c_i\}$  and product  $\{\bar{g}_i\}$ .

Let us first specify the power of the regulation-production-degradation noise  $\langle \mathcal{G}_i^2 \rangle$ , which has two contributions: super-Poissonian input noise  $\langle \mathcal{G}_{\text{TF},i}^2 \rangle$  originating from random arrival of regulatory transcription factors and Poissonian output noise  $\langle \mathcal{G}_{\text{PD},i}^2 \rangle$  from the production and degradation of the gene product. The input noise is a function of the transcription factor concentration and is well approximated by [11,55,56]

$$\langle \mathcal{G}_{\text{TF},i}^2 \rangle = \left[ \left( \frac{\partial r(c)}{\partial c} \right)^2 \frac{2c}{D_c l_c} \right]_{c=c_i}, \quad (13)$$

where  $D_c$  is the diffusion coefficient of the transcription factors,  $l_c$  the typical size of a transcription factor binding site, and  $c_i$  the local activator concentration. The above formula describes fluctuations of  $G_i$  in absolute units; to obtain the noise power  $\langle \hat{\mathcal{G}}_{\text{TF},i}^2 \rangle$  of fluctuations in the normalized output  $g_i$ ,  $\langle \mathcal{G}_{\text{TF},i}^2 \rangle$  has to be normalized as well by  $N_{\max}^2 = (r_{\max}\tau)^2$ :

$$\begin{aligned}\langle \hat{\mathcal{G}}_{\text{TF},i}^2 \rangle &= \frac{\langle \mathcal{G}_{\text{TF},i}^2 \rangle}{N_{\max}^2} \\ &= \frac{1}{N_{\max}\tau} \left[ \left( \frac{r_{\max} - r_b}{r_{\max}} \right)^2 \left( \frac{\partial f(c)}{\partial c} \right)^2 \frac{2c N_{\max}}{D_c l_c \tau} \right]_{c_i} \\ &= \frac{1}{N_{\max}\tau} \left[ (1 - \bar{g}_b)^2 \left( \frac{\partial f(c)}{\partial c} \right)^2 2cc_0 \right]_{c_i}.\end{aligned}\quad (14)$$

In the last step,  $c_0 \equiv N_{\max}/D_c l_c \tau$  defines a typical concentration scale for our problem; in the following we will measure concentrations in units of  $c_0$ .

The output noise in absolute units can be written as the sum of production and degradation rate, as for a regular birth-death process:

$$\langle \mathcal{G}_{\text{PD},i}^2 \rangle = r(c_i) + \frac{1}{\tau} \langle G_i \rangle. \quad (15)$$

Note that in general  $\langle G_i \rangle \neq r(c_i)\tau$  due to the diffusive coupling. Also note that bursty production (not considered in this work) could be easily incorporated into the model via a prefactor  $\phi$  (Fano factor) in the production term:  $r(c_i) \rightarrow \phi r(c_i)$ . In normalized units the production-degradation noise reads

$$\langle \hat{\mathcal{G}}_{\text{PD},i}^2 \rangle = \frac{\langle \mathcal{G}_{\text{PD},i}^2 \rangle}{N_{\max}^2} = \frac{1}{N_{\max}\tau} [\bar{g}_b + (1 - \bar{g}_b)f(c_i) + \bar{g}_i]. \quad (16)$$

In our model, diffusive hopping of a molecule is identical to its degradation in the volume of origin and simultaneous production of a molecule in the volume of arrival. The corresponding noise powers  $\mathcal{D}_{(i \rightarrow j)}^2$  thus are given by the rate

of particle loss through diffusive hopping from the volume of origin, which in the steady state is given by the mean copy number in that volume times the hopping rate, and can be normalized in the same way as the other noise powers:

$$\langle \mathcal{D}_{(i \rightarrow j)}^2 \rangle = \frac{D}{\delta^2} \langle G_i \rangle, \quad \langle \hat{\mathcal{D}}_{(i \rightarrow j)}^2 \rangle = \frac{\langle \mathcal{D}_{(i \rightarrow j)}^2 \rangle}{N_{\max}^2} = \frac{\Delta}{N_{\max} \tau} \bar{g}_i. \quad (17)$$

Adding up all contributions, we obtain for the (normalized) combined noise powers  $\mathcal{N}_{ij}^2$  and  $\mathcal{N}_{ii}^2$ ,

$$\mathcal{N}_{ij}^2 = -\langle \mathcal{D}_{(i \rightarrow j)}^2 \rangle - \langle \mathcal{D}_{(j \rightarrow i)}^2 \rangle = -\frac{\Delta}{N_{\max} \tau} (\bar{g}_i + \bar{g}_j), \quad (18)$$

$$\begin{aligned} \mathcal{N}_{ii}^2 &= \langle \mathcal{G}_{\text{PD},i}^2 \rangle + \langle \mathcal{G}_{\text{TF},i}^2 \rangle + \sum_{n \in N(i)} \langle \mathcal{D}_{(i \rightarrow n)}^2 + \mathcal{D}_{(n \rightarrow i)}^2 \rangle \\ &= \frac{1}{N_{\max} \tau} \left\{ \bar{g}_b + (1 - \bar{g}_b) f(c_i) + \bar{g}_i \right. \\ &\quad \left. + (1 - \bar{g}_b)^2 \left[ \left( \frac{\partial f(c)}{\partial c} \right)^2 2c c_0 \right]_{c_i} \right. \\ &\quad \left. + \Delta \sum_{n \in N(i)} (\bar{g}_i + \bar{g}_n) \right\}. \end{aligned} \quad (19)$$

#### D. Quantifying positional information

Building on our previous work [8,10–13], we quantify the amount of information that the noisy output signal  $g(x)$  carries about the position  $x$  using the mutual information  $I(x; g)$  between the input and output [57,58], a central quantity of Shannon's information theory [57]:

$$I(x; g) = \int dx P_x(x) \int dg P(g|x) \log_2 \left[ \frac{P(g|x)}{P_g(g)} \right]. \quad (20)$$

In the information-theoretic picture introduced here, the information channel is defined by a set of sampling units (cells or nuclei at a particular position) that implement the same gene regulatory network. These sampling units read out the input signal  $c$  with a distribution that is jointly determined by the shape of the concentration field  $c(x)$  and the spatial distribution of the sampling units  $P_x(x)$ . In response to these inputs, the regulatory networks locally express the output gene  $g$  at the corresponding level, producing an output distribution  $P_g(g)$  across the ensemble of sampling units.

The mutual information can be rewritten as a difference between the entropy of the output distribution  $S[P_g(g)] = -\int dg P_g(g) \log_2 P_g(g)$  and the average entropy of the conditional distribution of  $g$  at fixed position  $x$ :

$$I(x; g) = S[P_g(g)] - \int dx P_x(x) S[P(g|x)]. \quad (21)$$

When only the local distribution of outputs  $P(g|x)$  is known,  $P_g(g)$  can be straightforwardly constructed from  $P(g|x)$  and  $P_x(x)$ :

$$P_g(g) = \int dx P_x(x) P(g|x). \quad (22)$$

In our discrete model we assume uniform spacing of the sampling positions  $x_i \in [0, N_x \delta]$ , i.e.,  $P_x(x_i) = 1/N_x$ . Equations (21) and (22) then respectively become

$$\begin{aligned} I(x; g) &= - \int dg P_g(g) \log_2 P_g(g) \\ &\quad + \frac{1}{N_x} \sum_{i=1}^{N_x} \int dg P(g|x_i) \log_2 P(g|x_i), \end{aligned} \quad (23)$$

$$P_g(g) = \frac{1}{N_x} \sum_{i=1}^{N_x} P(g|x_i). \quad (24)$$

Therefore,  $I(x; g)$  is fully characterized by the conditional output distribution  $P(g|x_i)$ . Since here we only consider nonmultiplicative white noise,  $P(g|x_i)$  is Gaussian:

$$P(g|x_i) = \frac{1}{\sqrt{2\pi \sigma_g^2(x_i)}} \exp \left( -\frac{[g - \bar{g}(x_i)]^2}{2\sigma_g^2(x_i)} \right). \quad (25)$$

The information  $I(x; g)$  is therefore fully determined by the (local) means and variances of the conditional distributions  $P(g|x_i)$ .

In our previous work [8,10–13], we have computed the maximum achievable information transmission (channel capacity) by optimizing over the distribution of input signals, relying on the small noise approximation to make the problem analytically tractable. Here, in contrast, the spatial setting has led us to choose a uniform distribution of sampling points  $P_x(x)$ . While this relieves us of the need to use the small noise approximation, to fully specify the problem we still need to select the function  $c(x)$  that maps the input of the regulatory network to spatial positions. Motivated by the developmental example of the fruit fly, here we choose an exponential function [36,37]

$$c(x) = c_{\max} e^{-x/\lambda \gamma_0}, \quad (26)$$

where  $c_{\max}$  controls the maximum achievable input,  $\gamma_0 \equiv L/5 \equiv N_x \delta/5$  is the decay length of a typical exponential morphogen gradient relative to the system size  $L$  [36,37,59], and  $\lambda$  is a variable dimensionless scaling factor (the precise value of  $\gamma_0$  is insignificant; we could opt for any convenient length unit). This choice generates a family of input profiles parametrized by  $c_{\max}$  and  $\lambda$  and we will explore various settings for these parameters to maximize the positional information encoded by the expression level  $g$ .

We can now assemble the different components of the model. By inserting the regulation and noise terms [Eqs. (12), (18), and (19)] into the steady-state solutions of the coupled stochastic equations [Eqs. (9)–(11)] we obtain two coupled linear equation systems for mean expression levels  $\bar{g}_i$  and (co)variances  $C_{ij}$  and  $\sigma_i^2$  that we can solve given the selected boundary conditions. Using Eq. (23) we can then compute the mutual information  $I(x; g)$ . When computing the stochastic moments, we can vary the parameters of the input function, the regulation, and the spatial coupling and thus compute  $I(x; g)$  as a function of its key determinants. The baseline set of parameters that we used for numerical computation is presented in Appendix E.

### III. RESULTS

In order to elucidate how spatial coupling alters the capacity of encoding positional information in the downstream gene  $g$ , we optimize the mutual information  $I(x; g)$  over the parameters of our model: the dissociation constant  $K$  and the Hill coefficient of regulation  $H$ , the (normalized) diffusion constant of the gene product  $\Delta$ , and the length scale of the input gradient  $\lambda$ . As is known from previous work, the key parameter that qualitatively determines the shape of the optimal solutions is the ratio of the output to input noise, which is set by a dimensionless maximal input concentration  $C = c_{\max}/c_0$  [11]; in general,  $C \gg 1$  is the regime of dominant output noise, while  $C \ll 1$  is the regime of dominant input noise, as defined in Sec. II C. We thus explore the optimal solutions as a function of  $C$  in what follows. For computational efficiency we initially study the 1D model with the short-correlation assumption (SCA) and later compare to the 2D models with and without the SCA, finding only minor differences (see Sec. III D).

#### A. Spatial coupling can enhance information transmission

We first assess how introducing diffusive coupling changes the information capacity of the system in the space of regulatory parameters  $K$  and  $H$ . To establish the baseline for comparison, we start with the case without spatial coupling  $\Delta \simeq 0$ , setting  $C = 1$  and using the input gradient length  $\lambda = 1$ . Figure 2(a) shows the information plane for this case, i.e., the mutual information  $I(x; g)$  as a function of the Hill coefficient  $H$  and the activation threshold  $K$ . Consistently with our previous studies [10,11], the information plane displays a clear optimal choice of  $H$  and  $K$  that maximizes information transmission. The optimum results from a nontrivial compromise between evenly spreading the whole dynamic range of the output signal along the  $x$  axis, which ideally requires low Hill coefficients and half activation at the system center [ $x = L/2$ ,  $K = c(L/2)$ ], and a systemwide minimization of the noise; the latter generally favors activation at high input concentrations (to avoid large fluctuations at low  $c$ ) and thus higher  $K$ .

Figures 2(b) and 2(c) show the same information plane, but now at increasing values for the spatial coupling  $\Delta$ . The

first observation is that with nonzero  $\Delta$  information can be increased relative to the baseline at the optimal choice of  $H$  and  $K$ , but that the information drops again when the diffusive coupling is too strong, indicating that there exists a nontrivial optimal value for  $\Delta$  that maximizes information. The second observation relates to the concomitant change in the optimal parameters  $\{K^*, H^*\}$  as the diffusive coupling is increased. In particular, increasing the diffusion constant shifts the optimum towards higher  $H$  and lower  $K$ . The third observation is that strong diffusion also increases the capacity away from the optimum in the regime of high  $H$ , creating a flat plateau where the precise value of  $H$  is not crucial for attaining high information transmission.

All these observations can be understood intuitively: The increase in capacity for nonzero diffusion is due to the tradeoff between the ability of diffusive spatial averaging to reduce noise (thus increasing the transmission) and smoothing out the response profile (thus decreasing the transmission) [30,31]. The shift in optimal regulatory parameters is a direct consequence of these two effects: The detrimental effect of smoothing out the response profile can be partially compensated for by choosing higher optimal Hill coefficients, while noise reduction allows the optimal  $K$  to move towards lower concentrations to better utilize the dynamic range. The plateau in information at high diffusion results from the ability of the diffusion to flatten sharp, nearly steplike gene activation profiles at high  $H$  (limited to at most one bit of positional information for  $H \rightarrow \infty$  and  $\Delta \rightarrow 0$ ) into smooth spatial gradients that can convey more positional information.

#### B. Spatial coupling is most beneficial when input noise is dominant

After understanding the optimal behavior of information as a function of regulatory parameters  $H$  and  $K$  at fixed  $C = 1$ ,  $\lambda = 1$ , and  $\Delta$ , we systematically varied the diffusive coupling  $\Delta$  and the input gradient length scale  $\lambda$  for several different values of  $C$ . For each choice of  $(\Delta, \lambda)$  parameters, we separately mapped out the information as a function of  $H$  and  $K$  to find the optimal values  $(H^*, K^*)$  and the corresponding maximal information transmission.

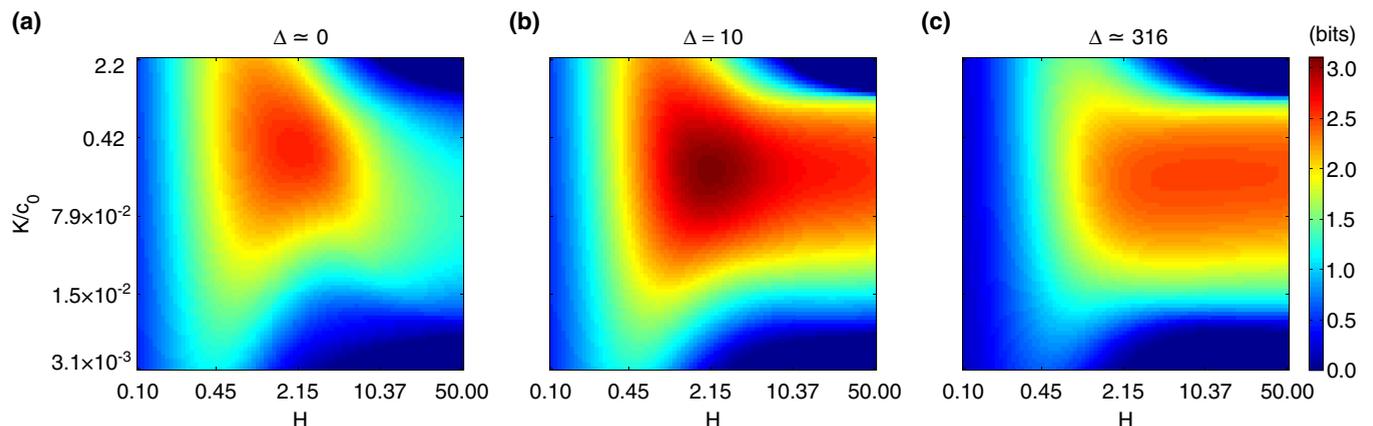


FIG. 2. (Color online) Information planes for different values of the spatial coupling  $\Delta$ . Shown is the mutual information  $I(x; g)$  as a function of the Hill coefficient  $H$  and the activation threshold  $K$  with (a) negligible spatial coupling  $\Delta \simeq 0$ , (b) intermediate spatial coupling  $\Delta = 10$ , and (c) very strong spatial coupling  $\Delta \simeq 316$ . The data shown are for the 1D model with the short-correlation assumption.

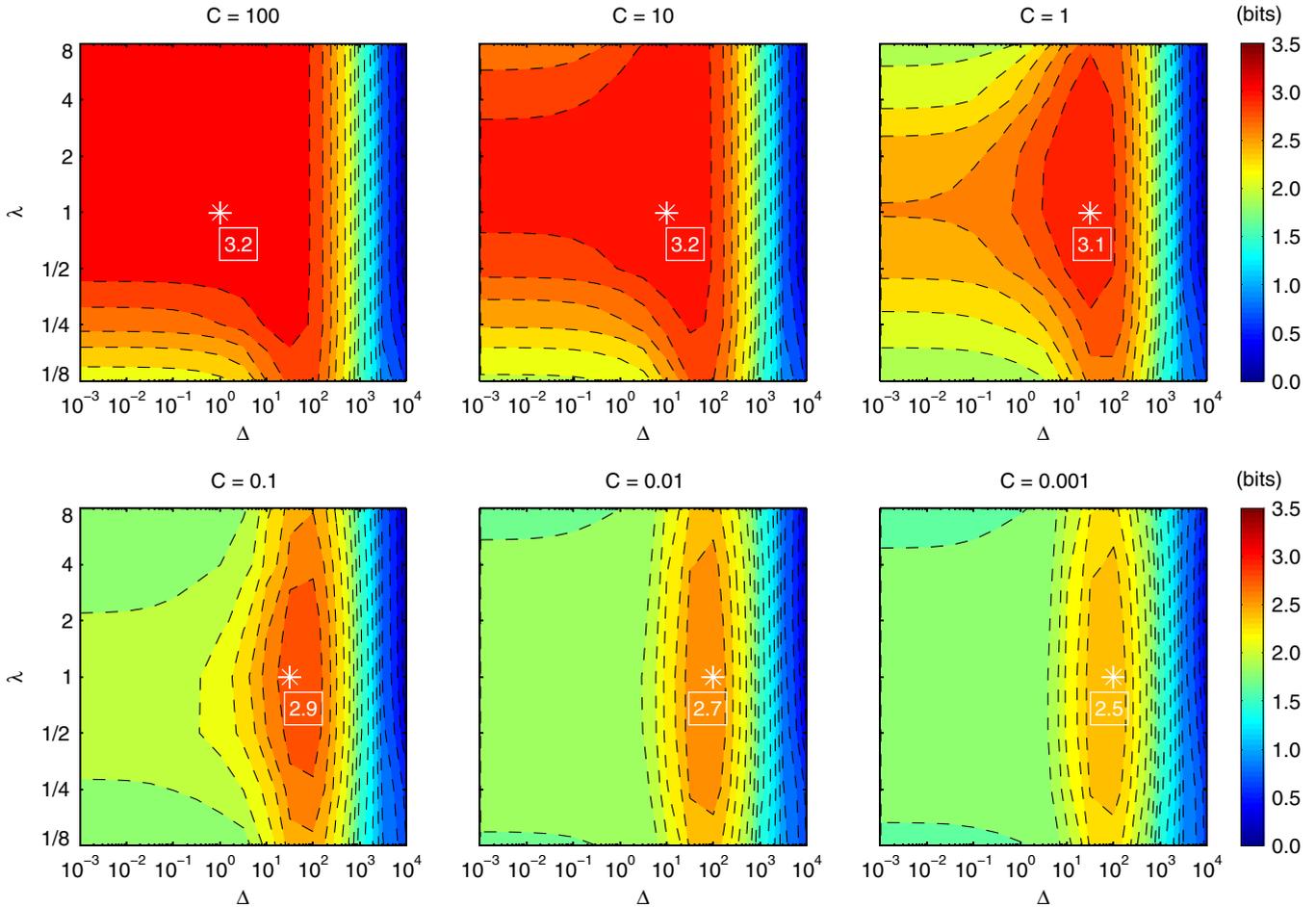


FIG. 3. (Color online) Information transmission as a function of the spatial coupling  $\Delta$  and the input gradient length  $\lambda$  for various values of maximal input concentration  $C$ . For comparison, the same scale is used for all contour maps. White asterisks mark the overall maximum of  $I(x; g)$  in each respective panel; white boxes show the corresponding optimal amount of transmitted information (in bits). Each data point was obtained by optimizing  $I(x; g)$  over the regulatory parameters  $H$  and  $K$  for the given set of  $C$ ,  $\Delta$ , and  $\lambda$ . The data shown are for the 1D model with the short-correlation assumption.

Figure 3 summarizes this exploration and shows the information capacity as a function of  $\Delta$  and  $\lambda$  for a series of six different  $C$  values. White asterisks mark the maximal capacity for the given  $C$ , i.e., the capacity for jointly optimal choice of  $H$ ,  $K$ ,  $\Delta$ , and  $\lambda$ .

The panel for  $C = 1$  demonstrates two beneficial effects of diffusion on information capacity: Increasing  $\Delta$  from very low values to  $\Delta \lesssim 100$  increases  $I(x; g)$  and simultaneously enlarges the range of  $\lambda$  over which high capacity can be reached. This effect is even more pronounced when  $C < 1$ : At very low  $C$  the beneficial values for the diffusion constant  $\Delta$  are strongly constrained to a narrow interval and the resulting capacity gain at optimal  $\Delta$  with respect to low  $\Delta$  increases markedly. In contrast, when  $C$  is high, diffusive coupling does not convey a large benefit in information transmission. This is due to the fact that diffusion can only remove super-Poissonian parts of the noise [30,31]. In our model, super-Poissonian noise is the input noise, which plays a minor role for  $C \gg 1$ , thereby limiting the potential role of diffusion. We note that super-Poissonian fluctuations could also occur on the output side (e.g., when gene expression is bursty and multiple protein copies are produced from each mRNA), in which case diffusive

coupling could be beneficial even for  $C \gg 1$ . As expected, for all  $C$ , the information capacity drops to zero at very high  $\Delta$  independently of all other parameters, because in that limit all output profiles become flat and thus uninformative.

We systematically explore how diffusion affects the optimal information transmission as a function of  $C$  in Fig. 4 by comparing the optimally coupled spatial model to a model where diffusion is set to zero. Because the information transmission is largely invariant to  $\lambda$  and for most values of  $C$  peaks in a very broad plateau at  $\lambda = 1$ , we set  $\lambda = 1$  for this comparison, while optimizing over all other parameters (separately for the spatially coupled system and the system with  $\Delta \simeq 0$ ). Clearly, in the low- $C$  regime optimal diffusive coupling can enhance information capacity by more than a bit, while for  $C \gtrsim 100$  the noise composition is strongly dominated by Poissonian output noise such that diffusion cannot further improve information throughput.

Instead of varying  $C$  by changing the maximal input concentration  $c_{\max}$ , we also vary  $C$  via  $N_{\max}$ , thus altering output noise at constant levels of input noise. In that case we observe the same behavior, i.e., significant capacity enhancement by diffusion in the low- $C$  regime, with the small difference that

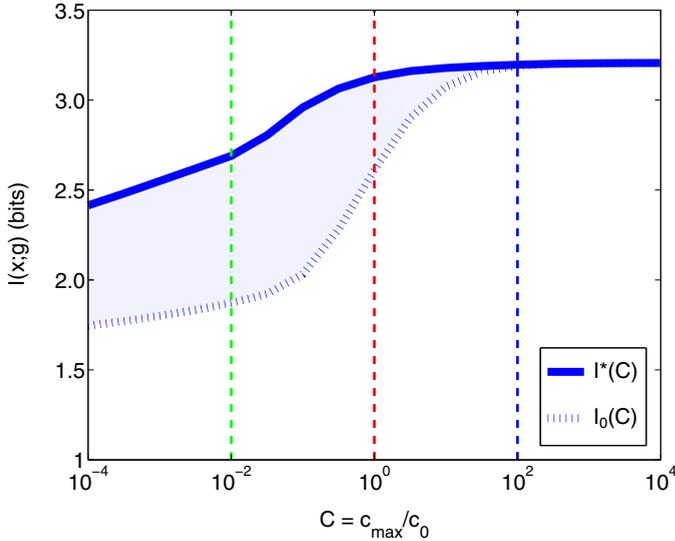


FIG. 4. (Color online) Information capacity for varying maximal input concentration  $C$ . Shown is the optimized information capacity as a function of  $C = c_{\max}/c_0$  with optimal diffusive coupling  $I^*(C)$  (solid blue line) and without diffusive coupling  $I_0(C)$  (dashed blue line). At small  $C$ , non-Poissonian input noise is dominant, while Poissonian output noise dominates for large  $C$ . The blue-shaded area depicts the maximal gain in information capacity from diffusive coupling: Spatial coupling enhances information capacity most efficiently when input noise dominates. The dashed vertical lines mark the  $C$  values corresponding to the three cases displayed in Fig. 5.

now the capacity increases with decreasing  $C \sim 1/N_{\max}$ , as presented in more detail in Appendix F.

### C. Spatial coupling enables a different regulatory strategy at high input noise

Spatial coupling affects both the noise level of the output gene and the shape of its output profile. We wondered how these two effects combine to affect the information transmission. Figure 5 shows the optimal mean output profile  $\bar{g}(\hat{x})$ , the underlying activation profile  $f(\hat{x})$ , and two measures of the noise in gene expression, the standard deviation  $\sigma^2(\hat{x})$  and the Fano factor  $\Phi(\hat{x}) = N_{\max}\hat{\sigma}^2(\hat{x})/\bar{g}(\hat{x})$ , as a function of position  $\hat{x} \equiv x/L \in [0, 1]$ . We plot these quantities for three values of  $C$ , representative of the input-noise-dominated regime ( $C = 0.01$ ), the regime in which input and output noise are approximately balanced ( $C = 1$ ), and the output-noise-dominated regime ( $C = 100$ ). In each plot we directly compare the system in which the diffusion constant is optimized along with all other parameters vs the system that is optimal at zero spatial coupling.

For  $C = 100$ , input noise levels are so low that the total noise is dominated by purely Poissonian output noise throughout the whole system [ $\Phi(\hat{x}) \simeq 1$ ]. Therefore, strong spatial coupling would fail to decrease the noise further and would merely fade out the mean output profile. Consequently, the optimal diffusion constant  $\Delta^* = 2$  is comparably low and the spatially coupled system only differs marginally from the uncoupled system.

For  $C = 1$  in the uncoupled system, input noise becomes more prominent as a fraction of the total noise such that  $\Phi(\hat{x}) > 1$  for  $\hat{x} \gtrsim 0.2$  [cyan colored line in Fig. 5(f)]; in line with this, the activation threshold is shifted towards higher input concentrations (i.e., smaller  $\hat{x}$ ) at the expense of reducing the dynamic range. In the coupled system the optimal diffusion constant is markedly higher ( $\Delta^* = 25$ ) than at  $C = 100$  and the resulting spatial averaging can almost completely remove the additional non-Poissonian noise contributions [ $\Phi(\hat{x}) \simeq 1$ ]. This allows the optimal activation threshold  $K$  to shift towards lower input concentrations in order to attain a similarly large dynamic range as for  $C = 100$ .

At  $C = 0.01$ , when input noise is exceedingly dominant, we observe the emergence of a qualitatively different regulatory strategy in the spatially coupled system. For both spatially coupled and uncoupled systems, low  $C$  favors sharp activation profiles with high Hill coefficients  $H$ . For the spatially uncoupled system, the optimal activation threshold  $K^*$  moves towards central positions. In contrast, in the optimal spatially coupled system, the optimal activation threshold  $K^*$  moves closer towards  $\hat{x} = 0$ , where input noise is smaller, and the optimal Hill coefficient makes the activation nearly steplike. Without spatial coupling, this strategy could not be optimal because it would be limited to at most one bit: In the region where the activation profile is flat, positions could not be discriminated based on the gene expression level  $g$ . Optimal diffusion can, however, smooth out this steplike activation to generate a graded gene expression profile and restore position discriminability at essentially all positions. Simultaneously, optimal diffusion transforms the noise in an interesting way, as can be seen in Figs. 5(h) and 5(i). The steplike activation curve suppresses the super-Poissonian input noise everywhere except for the narrow interval of  $x$  around the transition point where the super-Poisson component has a very sharp and tall peak; however, this can very effectively be removed by the strong diffusive spatial averaging. In sum, at low  $C$ , the optimal strategy is to use very sharp activation profiles since they act as a filter for input noise (which must go to zero at zero or saturated gene expression); the strong optimal diffusion then smoothes out the steplike activation into a graded response and removes much of the remaining input noise around the very localized transition region at  $c \approx K^*$ .

Usually, the effect of diffusion on information transmission is understood to be a tradeoff between the beneficial effect of noise averaging and detrimental effect of smoothing the profile, but here it seems that smoothing the steplike activation profile is also beneficial for the information flow. To check that this is indeed a qualitatively different optimal regulatory strategy, we performed a systematic study of the optimal regulatory parameters between  $C = 0.1$  and  $C = 0.01$ . We found that here two regulatory strategies compete, as evidenced by two local maxima in the  $(K, H)$  information planes for this range of  $C$  values: One favors relatively low Hill coefficients  $H^* \simeq 1$  and high activation thresholds  $K^*$  while the other favors high  $H^*$  and lower  $K^*$ . For  $C = 0.01$  and  $3 \lesssim \Delta \lesssim 10$  both regulatory strategies lead to almost equal information capacities, but for  $\Delta \gtrsim 10$  steep activation combined with fast diffusion is preferable. Figure 7 in Appendix G illustrates this phenomenon.

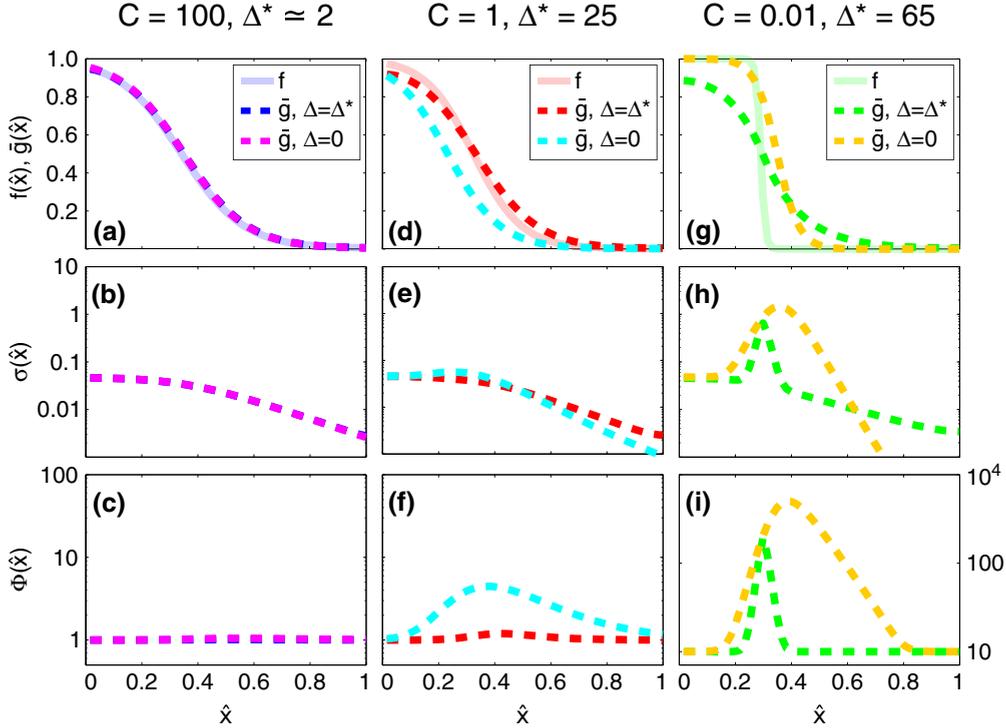


FIG. 5. (Color online) Comparison of optimal output profiles for different values of the maximal input concentration  $C$ : (a)–(c) dominant output noise  $C = 100$  (blue lines), (d)–(f) balanced noise  $C = 1$  (red lines), and (g)–(i) dominant input noise  $C = 0.01$  (green lines). For every value of  $C$ , the normalized mean output profile  $\bar{g}(\hat{x})$  (dashed lines) and the optimal activation profile  $f(\hat{x})$  (faint solid lines) are shown in the top row, the noise in gene expression is shown as the standard deviation  $\sigma(\hat{x})$  in the middle row, and the Fano factor  $\Phi(\hat{x}) = N_{\max}\sigma^2(\hat{x})/\bar{g}(\hat{x})$  is shown in the bottom row. All profiles are shown as a function of the normalized position  $\hat{x} = x/L$ . The corresponding quantities for the optimal spatially uncoupled ( $\Delta \simeq 0$ ) system are shown in complementary colors (see plot legends).

#### D. Full solution in 2D geometry differs only marginally from an approximate solution in one dimension

All results presented in the previous sections were obtained using the 1D model with the SCA. To assess how representative they are of the fully detailed model, we compared the 1D model with the SCA to the 2D model with the SCA and finally to the 2D model that retains more than only nearest-neighbor correlations (see Appendix B for the formulas in the first two cases and Appendix C for the latter case).

Figure 8 in Appendix H compares the information planes  $I(\Delta, \lambda)$  for the three cases at  $C = 1$ . The results suggest that the influence of lateral spatial coupling present in the 2D model is minor. Particularly in the range of spatial couplings  $\Delta$  around the information maximum, where the system is operating close to the hard bound set by irreducible Poissonian noise, the difference between information capacities in one and two dimensions is almost negligible. Inclusion of 2D couplings mainly increases  $I(x; g)$  values away from the optimal range of parameters, thus effectively enlarging the region of parameter space in which the system can attain close-to-optimal information transmission. This again is most prominent for the low- $C$  regime, as expected, but still limited to a small fraction ( $\lesssim 10\%$ ) of the total capacity (data not shown). The effect of incorporating longer-range correlations is minor as well: The information planes for the 2D model with and without the SCA are almost indistinguishable.

Overall, this demonstrates that the 1D model with the short-correlation assumption is a good approximation of the fully detailed model. In the given context, the strength of spatial coupling appears more relevant than its topology.

#### IV. DISCUSSION

We developed a generic framework to compute information flow through spatially coupled gene regulatory networks at the steady state. As the simplest example we considered how a spatially varying concentration field (input profile) is read out collectively by a regular 1D or 2D lattice of sampling units that are interacting by local diffusion of the gene product. A directly applicable example could be cells (or nuclei) in a developing multicellular organism that respond to the morphogen field by expressing developmental genes whose products can be exchanged between neighboring cells. We emphasize that our framework is generic in that effects due to spatial coupling in any chosen geometry can be worked out before assuming any particular gene regulatory function and writing down the relevant noise power spectra, as is evident from Eqs. (9)–(11). This makes the framework widely applicable to a broad range of biological problems in which spatially distributed information is collectively sensed by a set of discrete sampling units and encoded into a spatial output.

The theory presented here is a direct continuation of our previous work on quantifying information flow in gene regulatory networks at increasing levels of realism [10–14]. In our previous approaches, however, we were striving for analytical approximations that would permit computing the channel capacity, that is, performing analytical optimization over all possible distributions of input signals  $P_c(c)$ ; this led us to adopt the small noise approximation. Here, in contrast, we assumed a particular geometry [with a uniform distribution over sampling units  $P_x(x)$ ] and a particular functional form of the input concentration field  $c(x)$ ; the latter can depend on parameters that one can optimize, but we do not perform functional optimization over all possible  $c(x)$ . While these restrictions may appear stronger than in our previous work, they also allow us to move to a truly spatially discrete setup (that has a natural minimal spatial scale of a cell or nucleus, as is the case in reality) and permit us to relax the small noise approximation, which is no longer assumed in this work. We also note that a parametric choice of  $c(x)$  is not as restrictive as it might appear initially. First, one could choose a rich enough parametric form (basis set) for  $c(x)$  that ensures spatial smoothness but otherwise allows optimization in a full space of plausible profiles. Second, it is intuitively clear that monotonic functions encode positional information better than nonmonotonic or even constant functions, strongly narrowing the range of functional forms over which optimization should take place. Third, while we focused on exponential gradients that can be easily parametrized by only two parameters, exponential profiles actually are widespread throughout biology [23,24,42,59,60].

Several previous approaches assessed how diffusive coupling alters the precision of spatial protein patterns driven by spatially distributed inputs [30,31,61,62]. While similar in spirit to ours, these works employed measures of regulatory precision such as the output pattern steepness and sharpness, which, unlike mutual information, do not capture the problem in its full richness: These measures quantify precision only locally, on parts of the output signal, and moreover make implicit assumptions about which feature of the output (boundary steepness, boundary position, etc.) is informative. Information theory removes this arbitrariness [48] and permits extensions beyond the simple cases studied here. Related approaches [34,63], including an extended theory that accounts for the accessibility of information [64], have recently been suggested in similar biological contexts.

In our example application we studied how spatial coupling alters optimal information flow in a single-input–single-output gene regulatory network motivated by the bicoid-hunchback system, which is a part of the gap gene network in early fruit fly development [20,21]. The main outcome of this analysis is that diffusive coupling enhances information capacity by removing super-Poissonian components of the noise, in line with previous work [30,31]; this effect is large when input noise is non-negligible ( $C \leq 1$ ). When diffusion plays a large role in an optimal system, the activation functions can deviate substantially from the resulting output gene expression profiles. At the extreme, the activation functions can become steplike with very high Hill coefficients, while the gene expression profile still smoothly changes over its full dynamic range; this

strategy, optimal at low  $C$ , where both spatial noise averaging and profile smoothing due to diffusion act jointly to increase information, is qualitatively different from the  $C$  regime where profile smoothing is detrimental and trades off against noise averaging.

In our setup, super-Poissonian contributions to noise are fully accounted for by the input noise, but in reality this need not be the case. There are super-Poissonian contributions to noise that arise at the output, for instance, due to bursty production of proteins when the gene is activated. The simplest case is when mRNA expression is rate limiting, but then each mRNA can lead to a burst in the number of translated proteins. The theory can be simply extended to these cases by introducing a burst size into the relevant noise power spectra. Importantly, this would imply that diffusion can be effective at noise reduction (and thus beneficial for information transmission) also in the regime where  $C \geq 1$ . In support of this, recently another model linking mRNA expression and protein production with diffusion based on stochastic branching theory has shown that deviations from Poissonian statistics in typical biological conditions are expected only for rather large burst sizes ( $\gtrsim 50$ ) [65].

Recently, the field has made progress in detailed understanding of input-side noise in gene regulation due to stochastic diffusive arrivals of regulatory molecules to their binding sites [66,67]. This has led to a revision of the previously suggested functional form for the Berg-Purcell limit [68,69] that we use in Eq. (13). Furthermore, we have also taken the simplest (Hill-type) regulatory function as our model for gene regulation rather than picking a richer and potentially more realistic choice such as the Monod-Wyman-Changeux form. We will include these refinements in our model in the future, but there is no reason to expect that they could change any qualitative outcome of our analyses.

It is interesting to speculate about the predictive power of optimization-based approaches applied to biological systems. In neuroscience, the principle of efficient coding that states that neural sensory systems devote their limited resources to maximize the information flow, in bits, from the naturalistic stimuli into the spiking neural representation [70] has proven enormously successful since it was first suggested by Barlow [71]. In signaling and gene regulation, similar ideas are much more recent. On the other hand, the number of distinct genes or signaling proteins involved in the networks of interest is much smaller than the number of neurons; furthermore, molecular processes and the physical limits to sources of noise in gene regulation might be more easily understood than in neuroscience, where even a single neuron is a very complex object. It appears feasible that for small genetic networks the optimization problem would be tractable upon combining all phenomenology that until now we have analyzed separately: multiple interacting genes, driven by one or more inputs, potentially with arbitrary feedback, all major relevant sources of noise, spatial coupling, and readout constraints [10–14]. A major goal would also be the ability to relax the steady-state assumptions, by considering either readout at particular time points (out of steady state), or the information between full state trajectories [15–19]. Such a predictive theory, from which optimal networks could be derived mathematically, could then be realistically confronted with well-studied networks that can

be quantitatively measured, e.g., the gap gene network active during *Drosophila* development [20,21,39]. It is intriguing to think that even in the molecular world the need to pay for abstract bits of information—in the currency of time or energy—led nature to choose particular regulatory networks and perhaps poised them at special operating points [72].

### ACKNOWLEDGMENTS

The authors thank A. Mugler, I. Nemenman, T. Gregor, M. Tikhonov, P. R. ten Wolde, and K. Kaizu for fruitful discussions.

### APPENDIX A: CALCULATION OF STATIONARY MEANS AND (CO)VARIANCES

Considering Eq. (2) in steady state immediately yields

$$\begin{aligned} \langle F(g_i, c_i) \rangle &= \hat{r}_i - \frac{1}{\tau} \langle g_i \rangle \\ &= h \sum_{n_i \in N(i)} \langle g_i - g_{n_i} \rangle = 2dh \langle g_i \rangle - h \sum_{n_i \in N(i)} \langle g_{n_i} \rangle, \end{aligned} \quad (\text{A1})$$

where we abbreviate the normalized local production rate  $\hat{r}_i \equiv r_i / r_{\max} \tau$ . Equation (A1) simply reflects the steady-state balance between production (and degradation) and diffusive influx and outflux in volume  $i$ . Grouping  $\langle g_i \rangle$  terms on one side gives

$$\begin{aligned} \langle g_i \rangle &= \frac{\tau}{1 + 2dh\tau} \left( \hat{r}_i + h \sum_{n_i \in N(i)} \langle g_{n_i} \rangle \right) \Leftrightarrow \\ \bar{g}_i &= T \hat{r}_i + \Lambda^2 \sum_{n_i \in N(i)} \bar{g}_{n_i}, \end{aligned} \quad (\text{A2})$$

where in the last step we abbreviate  $\bar{g}_i \equiv \langle g_i \rangle$ ,  $T \equiv \tau / (1 + 2dh\tau)$ , and  $\Lambda^2 \equiv hT$ , as in the main text.

To simplify the steady-state expression for the covariances resulting from Eq. (3) it is instructive to treat its different parts separately. For the terms correlating the fluctuations in volume  $i$  with the production-degradation process in volume  $j$  we have (note that  $\langle \delta g_i \rangle = 0$  by construction)

$$\begin{aligned} \langle \delta g_i F(g_j, c_j) \rangle &= \langle \delta g_i \rangle \hat{r}_j - \frac{1}{\tau} \langle \delta g_i \delta g_j \rangle \\ &= -\frac{1}{\tau} \langle \delta g_i ((g_j) + \delta g_j) \rangle = -\frac{1}{\tau} \langle \delta g_i \delta g_j \rangle. \end{aligned} \quad (\text{A3})$$

Hence,

$$\langle \delta g_i F(g_j, c_j) \rangle + \langle \delta g_j F(g_i, c_i) \rangle = -\frac{2}{\tau} \langle \delta g_i \delta g_j \rangle \quad (\text{A4})$$

because  $\langle \delta g_i \delta g_j \rangle \equiv \langle \delta g_j \delta g_i \rangle$ . Similarly, we can rewrite the term that correlates  $\delta g_i$  with the diffusive neighbor fluxes of  $g_j$ ,

$$\begin{aligned} &h \left\langle \delta g_i \sum_{n_j \in N(j)} (g_{n_j} - g_j) \right\rangle \\ &= h \left\langle \delta g_i \sum_{n_j \in N(j)} ((g_{n_j}) + \delta g_{n_j} - \langle g_j \rangle - \delta g_j) \right\rangle \end{aligned}$$

$$\begin{aligned} &= h \left( \sum_{n_j \in N(j)} \langle \delta g_i \rangle \langle g_{n_j} \rangle + \langle \delta g_i \delta g_{n_j} \rangle \right) \\ &\quad - 2dh (\langle \delta g_i \rangle \langle g_j \rangle - \langle \delta g_i \delta g_j \rangle) \\ &= h \left( \sum_{n_j \in N(j)} \langle \delta g_i \delta g_{n_j} \rangle \right) - 2dh \langle \delta g_i \delta g_j \rangle, \end{aligned} \quad (\text{A5})$$

and analogously for the term in which  $i$  and  $j$  are exchanged. Reinserting (A4) and (A5) into the right-hand side of Eq. (3), collecting covariance terms  $\langle \delta g_i \delta g_j \rangle$  on the left-hand side, and multiplying by  $\tau/2$  we obtain

$$\begin{aligned} &(1 + 2dh\tau) \langle \delta g_i \delta g_j \rangle \\ &= \frac{h\tau}{2} \left( \sum_{n_j \in N(j)} \langle \delta g_i \delta g_{n_j} \rangle + \sum_{n_i \in N(i)} \langle \delta g_j \delta g_{n_i} \rangle \right) \\ &\quad + \frac{\tau}{2} \sum_k \langle \Gamma_{ik} \Gamma_{jk} \rangle \\ \Leftrightarrow C_{ij} &= \frac{\Lambda^2}{2} \left( \sum_{n_j \in N(j)} C_{in_j} + \sum_{n_i \in N(i)} C_{jn_i} \right) \\ &\quad + \frac{T}{2} \sum_k \langle \Gamma_{ik} \Gamma_{jk} \rangle, \end{aligned} \quad (\text{A6})$$

where  $C_{ij} \equiv \langle \delta g_i \delta g_j \rangle$ . This equation couples the covariance  $C_{ij}$  to the covariances  $C_{in_j}$  and  $C_{jn_i}$ ; these represent the correlations between the protein number in volume  $i$  and the neighbor volumes  $n_j$  of  $j$ , and the analogous quantity with  $i$  and  $j$  exchanged, respectively. Hence, even if  $i$  and  $j$  are nearest neighbors on the lattice, the expressions summing  $C_{in_j}$  and  $C_{jn_i}$  over  $n_j$  and  $n_i$ , respectively, will contain correlations between next-nearest-neighbor volumes. While the equation system defined by Eq. (A6) can be solved for the whole set of covariances  $C_{ij}$  upon imposing suitable boundary conditions, this can be numerically expensive for larger spatial lattices and large parameter sweeps as part of optimization runs. In this work the quantity of interest is the variance  $C_{ii}$  in volume  $i$  such that the calculation of longer-range correlations is not strictly required.

Considering the cases  $i = j$  and  $v \in N(i)$  ( $v$  being one of the nearest neighbors of  $i$ ) in Eq. (A6) separately reveals the following interdependence between the variance  $\sigma_i^2 \equiv C_{ii}$  and the nearest-neighbor covariance  $C_{iv}$ :

$$\sigma_i^2 = \Lambda^2 \sum_{v \in N(i)} C_{iv} + \frac{T}{2} \sum_k \langle \Gamma_{ik}^2 \rangle, \quad (\text{A7})$$

$$\begin{aligned} C_{iv} &= \frac{\Lambda^2}{2} \left( \sum_{n_v \in N(v)} C_{in_v} + \sum_{n_i \in N(i)} C_{vn_i} \right) \\ &\quad + \frac{T}{2} \sum_k \langle \Gamma_{ik} \Gamma_{vk} \rangle. \end{aligned} \quad (\text{A8})$$

Next-nearest correlations now only appear in Eq. (A8), which can be significantly simplified by an approximation that we call the short-correlation assumption: If the next-nearest-

neighbor covariances are assumed to be small compared to the nearest-neighbor covariances and single-point variances, we can ignore them and set

$$\begin{aligned} C_{in_v} &= 0 & \forall n_v \neq i, \\ C_{vn_i} &= 0 & \forall n_i \neq v \end{aligned} \quad (\text{A9})$$

because then the only neighbor of  $v$  that has some significant correlation with  $i$  is  $i$  itself and vice versa. In that case, the only remaining terms of the sums in Eq. (A8) are  $C_{ii} = \sigma_i^2$  and  $C_{vv} = \sigma_v^2$ , respectively:

$$C_{iv} = \frac{\Lambda^2}{2} (\sigma_i^2 + \sigma_v^2) + \frac{T}{2} \sum_k \langle \Gamma_{ik} \Gamma_{vk} \rangle. \quad (\text{A10})$$

Specifying further the noise powers  $\sum_k \langle \Gamma_{ik}^2 \rangle$  in (A7) and  $\sum_k \langle \Gamma_{ik} \Gamma_{vk} \rangle$  in (A10) as described in Sec. II B yields formulas (10) and (11).

### APPENDIX B: SIMPLIFIED VARIANCE FORMULAS

In our framework, computation of the mutual information  $I(x; g)$  only requires the means  $\bar{g}_i$  and variances  $\sigma_i^2$ . This enables us to reduce the number of coupled equations to be solved by direct insertion of (11) into (10), thus eliminating  $C_{ij}$ . For the 1D model with the SCA, after some algebraic steps this yields

$$\begin{aligned} \sigma_i^2 &= \frac{\Delta^2}{(1 + 2\Delta)^2 - \Delta^2} \frac{1}{2} (\sigma_{i-1}^2 + \sigma_{i+1}^2) \\ &+ \frac{1}{N_{\max}} \left\{ \frac{1 + 2\Delta}{(1 + 2\Delta)^2 - \Delta^2} \left\{ \frac{\bar{g}_b + (1 - \bar{g}_b) f(c_i) + \bar{g}_i}{2} \right. \right. \\ &+ \left. \left. \left[ (1 - \bar{g}_b)^2 \left( \frac{\partial f}{\partial c} \right)^2 c c_0 \right]_{c_i} \right\} \right. \\ &+ \left. \frac{\Delta}{1 + 3\Delta} \frac{1}{2} (\bar{g}_{i-1} + 2\bar{g}_i + \bar{g}_{i+1}) \right\}, \end{aligned} \quad (\text{B1})$$

where we have rewritten prefactor combinations containing  $T$  and  $\Lambda^2$  in terms of the spatial coupling  $\Delta$ . As explained in the main text (Sec. II B),  $\bar{g}_b$  is the basal expression level of the protein in normalized units.

The analogous formula for the 2D model with the SCA reads

$$\begin{aligned} \sigma_{(ij)}^2 &= \frac{\Delta^2}{(1 + 4\Delta)^2 - 2\Delta^2} \frac{1}{2} \sum_{n_{(ij)}} \sigma_{n_{(ij)}}^2 \\ &+ \frac{1}{N_{\max}} \left\{ \frac{1 + 4\Delta}{(1 + 4\Delta)^2 - 2\Delta^2} \left\{ \frac{\bar{g}_b + (1 - \bar{g}_b) f(c_{(ij)}) + \bar{g}_{(ij)}}{2} \right. \right. \\ &+ \left. \left. \left[ (1 - \bar{g}_b)^2 \left( \frac{\partial f}{\partial c} \right)^2 c c_0 \right]_{c_{(ij)}} \right\} \right. \\ &+ \left. \frac{\Delta(1 + 3\Delta)}{(1 + 4\Delta)^2 - 2\Delta^2} \frac{1}{2} \left( 4\bar{g}_{(ij)} + \sum_{n_{(ij)}} \bar{g}_{n_{(ij)}} \right) \right\}, \end{aligned} \quad (\text{B2})$$

where we indicate volumes of the two-dimensional lattice with a 2D index  $(ij)$  and the  $n_{(ij)}$  sums run over the four nearest-neighbor volumes of  $(ij)$ . Note the different form of the

prefactors containing the spatial coupling  $\Delta$  when compared to the 1D formula.

### APPENDIX C: FULL SOLUTION WITHOUT THE SHORT-CORRELATION ASSUMPTION

For completeness, here we also state the formula defining the linear system for the coupled covariances in the full two-dimensional model without the short-correlation assumption, again indicating volumes by the two-dimensional index  $(ij)$ :

$$\begin{aligned} C_{(ij)(kl)} &= \frac{1}{2} \frac{1}{1 + 4\Delta} \{ \hat{\mathcal{N}}_{(ij)(kl)}^2 + \Delta [C_{(ij)(k-1,l)} + C_{(ij)(k+1,l)} \\ &+ C_{(ij)(k,l-1)} + C_{(ij)(k,l+1)} + C_{(i-1,j)(kl)} \\ &+ C_{(i+1,j)(kl)} + C_{(i,j-1)(kl)} + C_{(i,j+1)(kl)}] \}, \end{aligned} \quad (\text{C1})$$

where the normalized noise term  $\hat{\mathcal{N}}_{(ij)(kl)}^2$  is only nonzero if the volumes indicated by  $(ij)$  and  $(kl)$  are identical, i.e.,  $(ij) \equiv (kl)$ , or nearest neighbors, i.e.,  $(kl) \in N((ij))$ , and takes one of the following forms, respectively:

$$\begin{aligned} \hat{\mathcal{N}}_{(ij)(kl)}^2 &= \hat{\mathcal{N}}_{(ij)(ij)}^2 = \hat{\mathcal{G}}_{(ij)}^2 + \sum_{n_{(ij)}} (\hat{\mathcal{D}}_{[(ij) \rightarrow n_{(ij)}]}^2 + \hat{\mathcal{D}}_{[n_{(ij)} \rightarrow (ij)]}^2) \\ &\quad \text{for } (ij) \equiv (kl), \\ \hat{\mathcal{N}}_{(ij)(kl)}^2 &= -(\hat{\mathcal{D}}_{[(ij) \rightarrow (kl)]}^2 + \hat{\mathcal{D}}_{[(kl) \rightarrow (ij)]}^2) \quad \text{for } (kl) \in N((ij)), \\ \hat{\mathcal{N}}_{(ij)(kl)}^2 &\equiv 0 \quad \text{otherwise.} \end{aligned} \quad (\text{C2})$$

Here  $n_{(ij)}$  runs over the four nearest neighbors of volume  $(ij)$ . The normalized noise powers  $\hat{\mathcal{G}}_{(ij)}^2$  and  $\hat{\mathcal{D}}_{[(ij) \rightarrow (kl)]}^2$  derive from the expressions presented in Sec. II C:

$$\begin{aligned} \hat{\mathcal{G}}_{(ij)}^2 &= \frac{1}{N_{\max}} \left\{ \bar{g}_b + (1 - \bar{g}_b) f(c_{(ij)}) + \bar{g}_{(ij)} \right. \\ &+ \left. (1 - \bar{g}_b)^2 \left[ \left( \frac{\partial f}{\partial c} \right)^2 2c c_0 \right]_{c_{(ij)}} \right\}, \end{aligned} \quad (\text{C3})$$

$$\hat{\mathcal{D}}_{[(ij) \rightarrow (kl)]}^2 = \frac{\Delta}{N_{\max}} \bar{g}_{(ij)}.$$

Equation (C1) in principle allows for calculation of covariances between volumes that are arbitrarily far apart. In practice, because of the finiteness of the lattice, the spatial correlations still have to be truncated at a certain distance and boundary conditions applied in order to obtain a closed system that consists of as many equations as variables.

### APPENDIX D: SMALL BASAL EXPRESSION REGULARIZES THE THEORY FOR VERY STEEP OUTPUT GENE EXPRESSION PROFILES

In the regime of steep activation curves and low diffusion constants the output profiles in our model become almost step-like, with anterior and posterior activation levels approaching 1 and 0, respectively. However, for any finite value of the Hill coefficient  $H$ , posterior expression levels never precisely reach zero, but rather form a tail of exponentially vanishing output values (as low as  $\sim 10^{-100}$  copies). As expected, our continuous approximation to gene expression levels must break down in this regime. This is indicated by spurious

TABLE I. Standard parameter values of the model.

Quantity	Symbol	Value
lattice spacing	$\delta$	$8.33 \mu\text{m}$
No. of nuclei in the $x$ direction	$N_x$	60
resulting system length	$L$	$500 \mu\text{m}$
internal activator diffusion constant	$D_c$	$3.16 \mu\text{m}^2/\text{s}$
activator binding site length	$l_c$	$0.01 \mu\text{m}$
product protein lifetime	$\tau$	240 s
basal expression rate	$r_b$	$0.01/\tau$
standard maximum mean product copy No.	$N_{\max}^0$	444
resulting typical concentration	$c_0$	$58.5 \mu\text{m}^{-3}$ ( $\approx 35 \text{ nM}$ )
resulting typical diffusion constant	$D_0$	$0.29 \mu\text{m}^2/\text{s}$
standard input gradient length	$\gamma_0$	$100 \mu\text{m}$
standard input gradient amplitude	$c_{\max}$	$c_0$

contributions to positional information arising from such small physically unrealizable levels of expression that nevertheless remain formally distinguishable between neighboring discrete positions in the lattice, given Gaussian noise with Poisson variance. We consider these contributions an unphysical artifact of our continuous theory, which is easily eliminated by either thresholding and setting very small nonzero expression levels (corresponding to less than one copy) to zero or by introducing a small basal expression level. We opted for the second solution because leaky gene expression is an experimentally verified observation [73], with a small basal rate  $r_b = 0.01/\tau$ . The precise value of  $r_b$  is unimportant, as long as it is small compared to the full dynamic range, yet significantly different from zero.

#### APPENDIX E: STANDARD PARAMETERS

In our derivations most system parameters can be combined into two natural scales, a typical concentration  $c_0$  (cf. Sec. IIC) and a typical diffusion constant  $D_0$  (cf. Sec. IIB), and we measure concentrations and diffusion in these scales throughout our theory. To obtain numerical solutions, however, concrete numbers have to be assigned to these quantities.

Since our example application represents the bicoid-hunchback system in early *Drosophila* development, we opted for the following choice for the baseline values of the parameters in Table I. First we chose lattice parameters that roughly correspond to the geometry of the *Drosophila* syncytium [37,38]; in particular, this defines the lattice spacing  $\delta$ . We then chose a typical value for the product protein lifetime  $\tau$ , which defines the typical diffusion scale  $D_0 = \delta^2/\tau$ .

To set the value for  $c_0$ , we took advantage of the fact that in the bicoid-hunchback system the activator (i.e., Bicoid) concentration at midembryo has been measured experimentally [36,37]. This lead us to set  $c_0$  equal to the extrapolated maximal value of the measured *in vivo* gradient at its source ( $x = 0$ ). In other words, at  $C = c_{\max}/c_0 = 1$  and  $\lambda = 1$ , the input function  $c(x)$  in our model corresponds to the experimentally measured Bicoid gradient. Finally, to define a standard value for the maximal mean output  $N_{\max}$ , which is a key determinant of the noise powers and variances [cf. Eqs. (18), (19), (B1), (B2), and (C3)] but to-date unknown, we chose typical values for the internal activator diffusion constant

$D_c$  and the binding site length  $l_c$ ; from this, we calculated the standard value  $N_{\max}^0 = c_0 D_c l_c \tau$ . When varying  $C$  away from the baseline setting, we changed  $c_{\max}$  either by holding  $N_{\max}$  (and thus  $c_0$ ) constant or by varying  $N_{\max}$  and keeping  $c_{\max}$  unchanged, as described in Appendix F. Table I demonstrates that all baseline parameter values are well within a biologically realistic regime.

#### APPENDIX F: VARYING $C$ VIA $N_{\max}$

In our model the noise-type ratio  $C = c_{\max}/c_0 = (c_{\max}/N_{\max})D_c l_c \tau$  can be varied in multiple ways. In addition

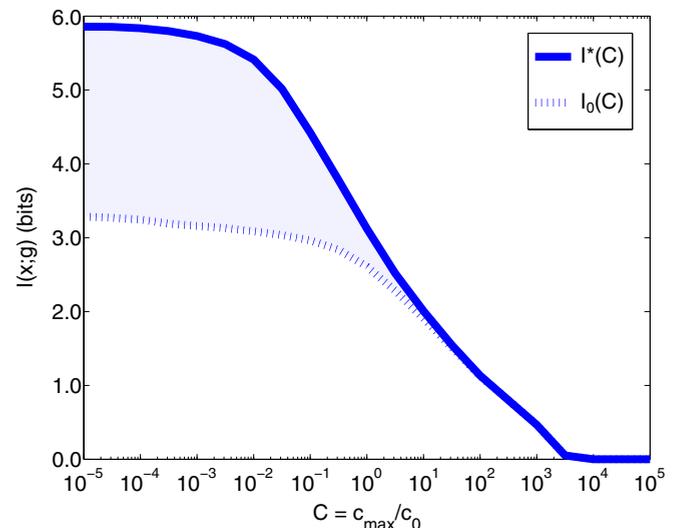


FIG. 6. (Color online) Information capacity for varying noise type ratios, varied via  $N_{\max}$ . Shown is the optimized information capacity as a function of the noise type ratio  $C = c_{\max}/c_0$  with optimal diffusive coupling  $I^*(C)$  (solid blue line) and without diffusive coupling  $I_0(C)$  (dashed blue line);  $C$  was varied via  $N_{\max}$ . At small  $C$  values, non-Poissonian input noise is dominant, while Poissonian output noise dominates for large  $C$ . The blue-shaded area depicts the maximal gain in information capacity from diffusive coupling.

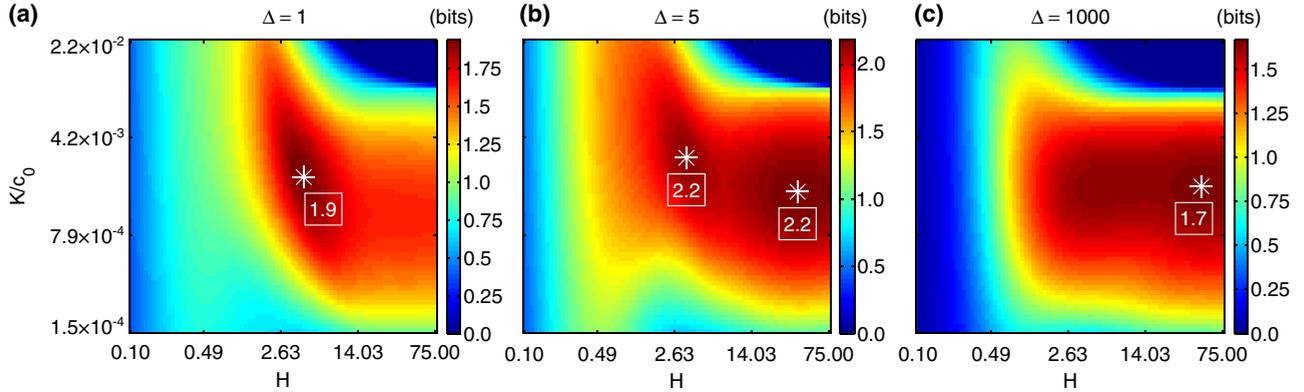


FIG. 7. (Color online) Emergence of a second maximum in the information plane. Here we plot for  $C = 0.01$  and standard input gradient length, i.e.,  $\lambda = 1$ , the positional information  $I(x; g)$  as a function of the regulatory parameters  $H$  and  $K$  for increasing strength of spatial coupling: (a)  $\Delta = 1$ , (b)  $\Delta = 5$ , and (c)  $\Delta = 1000$ . White asterisks mark local maxima, white boxes show the corresponding amount of information in bits. The data shown are for the 2D model with the SCA.

to altering the importance of non-Poissonian input noise via  $c_{\max}$ , we also studied the case in which the contribution of Poissonian output noise is varied via  $N_{\max}$  while  $c_{\max}$  is held constant.

The main difference from the case in which  $c_{\max}$  is scaled is that now information capacities decrease with increasing  $C$ , as this means decreasing  $N_{\max}$  and thus enhancing output noise (see Fig. 6). The highest information capacity then is attained in the limit  $C \rightarrow 0$ , i.e.,  $N_{\max} \rightarrow \infty$ . In the uncoupled system, for  $C \rightarrow 0$  the information  $I(x; g)$  approaches a value dictated by the amount of (in this case) irreducible input noise, set by  $c_{\max}$ . With spatial coupling, optimal  $I(x; g)$  values in the low- $C$  regime are markedly higher, in accordance with the finding that spatial averaging can only enhance information capacity when input noise is dominant. As expected, for  $C \rightarrow 0$ , i.e., negligible output noise, and sufficiently strong spatial coupling, i.e., strongly attenuated input noise, the capacity saturates at the hard bound of the noise-free limit given by the finite number of sampling points  $N_x$  along the  $x$  axis,  $I_{N_x}^{\max} = \log_2(N_x)$ . In the opposite limit  $C \rightarrow \infty$ ,  $I(x; g)$  drops to zero both with and without spatial coupling

when  $N_{\max}$  gets below the average basal expression level  $r_b \tau = 0.01$ .

#### APPENDIX G: DOUBLE MAXIMUM

In Fig. 7 we show three different information planes for the noise-type ratio  $C = 0.01$  and increasing values of the spatial coupling  $\Delta$  ( $\Delta = 1, 5$ , and  $1000$ ). The heat maps demonstrate the emergence of distinct optimal regulatory strategies for  $\Delta > 1$ : For sufficiently strong diffusion ( $\Delta = 5$ ), very steep activation curves ( $H \gg 1$ ) at lower activation thresholds  $K$  result in similar information transmission as less steep activation curves ( $H \simeq 1$ ) at slightly higher  $K$ . For  $\Delta = 1000$ , steep activation performs better than the other strategy. This effect is seen most clearly in the 2D model (with the SCA), but also appears in the 1D model at slightly different values of  $C$  and  $\Delta$ .

#### APPENDIX H: MODEL COMPARISON

In Fig. 8 we demonstrate how including increasing detail affects our results for a paradigmatic case ( $C = 1, \lambda = 1$ ).

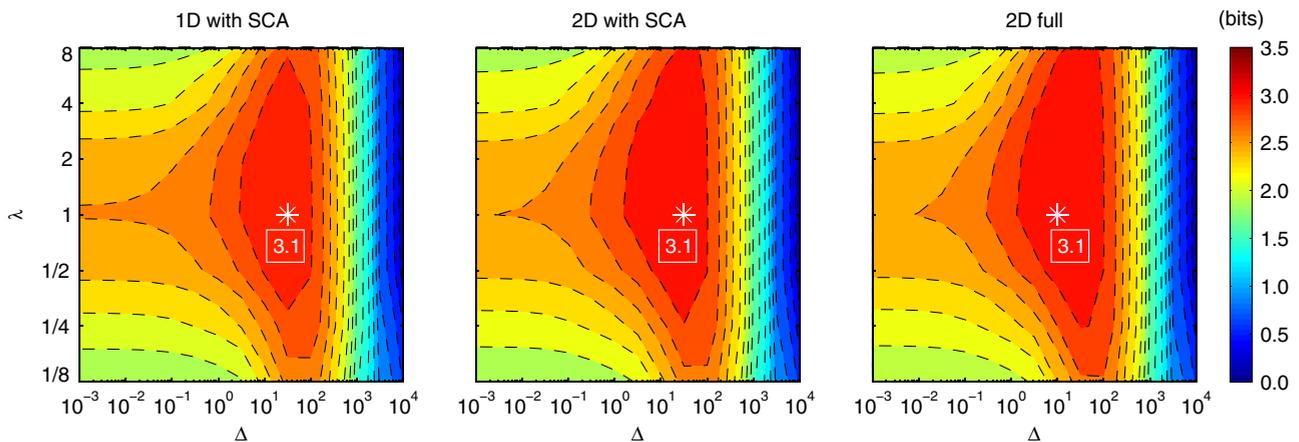


FIG. 8. (Color online) Comparison of  $I(x; g)$  for increasingly detailed versions of the spatial-stochastic model at  $C = 1$ . The reference plot for the 1D model with the SCA (left) is identical to the information plane for  $C = 1$  in Fig. 3. The same information plane is shown for the 2D model with the SCA (middle) and for the full 2D model that retains next-nearest-neighbor correlations (right).

The different panels show the same information plane for the 1D model with the SCA, the 2D model with the SCA, and

the 2D model without the SCA, which retains longer-range correlations with next-nearest-neighbor volumes.

- 
- [1] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- [2] L. S. Tsimring, *Rep. Prog. Phys.* **77**, 026601 (2014).
- [3] W.-C. Lo, S. Zhou, F. Y.-M. Wan, A. D. Lander, and Q. Nie, *J. R. Soc. Interface* **12**, 20141041 (2015).
- [4] L. Zhang, K. Radtke, L. Zheng, A. Q. Cai, T. F. Schilling, and Q. Nie, *Mol. Syst. Biol.* **8**, 613 (2012).
- [5] A. Eldar, B.-Z. Shilo, and N. Barkai, *Curr. Opin. Genet. Dev.* **14**, 435 (2004).
- [6] A. Eldar, R. Dorfman, D. Weiss, H. Ashe, B.-Z. Shilo, and N. Barkai, *Nature (London)* **419**, 304 (2002).
- [7] G. Tkačik and A. M. Walczak, *J. Phys.: Condens. Matter* **23**, 153102 (2011).
- [8] G. Tkačik, C. G. Callan, and W. Bialek, *Proc. Natl. Acad. Sci. USA* **105**, 12265 (2008).
- [9] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, *Proc. Natl. Acad. Sci. USA* **110**, 16301 (2013).
- [10] G. Tkačik, C. G. Callan, and W. Bialek, *Phys. Rev. E* **78**, 011910 (2008).
- [11] G. Tkačik, A. M. Walczak, and W. Bialek, *Phys. Rev. E* **80**, 031920 (2009).
- [12] A. M. Walczak, G. Tkačik, and W. Bialek, *Phys. Rev. E* **81**, 041905 (2010).
- [13] G. Tkačik, A. M. Walczak, and W. Bialek, *Phys. Rev. E* **85**, 041903 (2012).
- [14] G. Rieckh and G. Tkačik, *Biophys. J.* **106**, 1194 (2014).
- [15] F. Tostevin and P. R. ten Wolde, *Phys. Rev. Lett.* **102**, 218101 (2009).
- [16] F. Tostevin and P. R. ten Wolde, *Phys. Rev. E* **81**, 061917 (2010).
- [17] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde, *Phys. Rev. E* **82**, 031914 (2010).
- [18] A. Mugler, A. M. Walczak, and C. H. Wiggins, *Phys. Rev. Lett.* **105**, 058101 (2010).
- [19] J. Selimkhanov, B. Taylor, J. Yao, A. Pilko, J. Albeck, A. Hoffmann, L. Tsimring, and R. Wollman, *Science* **346**, 1370 (2014).
- [20] J. Jaeger, *Cell Mol. Life Sci.* **68**, 243 (2011).
- [21] J. Jaeger, Manu, and J. Reinitz, *Curr. Opin. Genet. Dev.* **22**, 533 (2012).
- [22] C. A. Rushlow and S. Y. Shvartsman, *Curr. Opin. Genet. Syst. Biol.* **22**, 542 (2012).
- [23] S. Restrepo, J. Zartman, and K. Basler, *Curr. Biol.* **24**, R245 (2014).
- [24] O. Wartlick, A. Kicheva, and M. González-Gaitán, *Cold Spring Harb. Perspect. Biol.* **1**, a001255 (2009).
- [25] H. Meinhardt, *Models of Biological Pattern Formation* (Academic, London, 1982).
- [26] T. Gregor, K. Fujimoto, N. Masaki, and S. Sawai, *Science* **328**, 1021 (2010).
- [27] K. Kamino, K. Fujimoto, and S. Sawai, *Dev. Growth Differ.* **53**, 503 (2011).
- [28] B. Sun, J. Lembong, V. Normand, M. Rogers, and H. A. Stone, *Proc. Natl. Acad. Sci. USA* **109**, 7753 (2012).
- [29] B. Sun, G. Duclos, and H. A. Stone, *Phys. Rev. Lett.* **110**, 158103 (2013).
- [30] T. Erdmann, M. Howard, and P. R. ten Wolde, *Phys. Rev. Lett.* **103**, 258101 (2009).
- [31] T. R. Sokolowski, T. Erdmann, and P. R. ten Wolde, *PLoS Comput. Biol.* **8**, e1002654 (2012).
- [32] S. C. Little, M. Tikhonov, and T. Gregor, *Cell* **154**, 789 (2013).
- [33] H. Garcia, M. Tikhonov, A. Lin, and T. Gregor, *Curr. Biol.* **23**, 2140 (2013).
- [34] A. Mugler, M. D. Brennan, A. Levchenko, and I. Nemenman, Presentation at the Eighth q-bio Conference, Sante Fe, 2014 (unpublished), and private communication.
- [35] M. Tikhonov, S. C. Little, T. Gregor, and W. Bialek, Presentation at the Eighth q-bio Conference, Sante Fe, 2014 (Ref. [34]), and private communication.
- [36] B. Houchmandzadeh, E. Wieschaus, and S. Leibler, *Nature (London)* **415**, 798 (2002).
- [37] T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank, *Cell* **130**, 141 (2007).
- [38] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, *Cell* **130**, 153 (2007).
- [39] J. O. Dubuis, R. Samanta, and T. Gregor, *Mol. Syst. Biol.* **9**, 639 (2013).
- [40] M. D. Petkova, S. C. Little, F. Liu, and T. Gregor, *Curr. Biol.* **24**, 1283 (2014).
- [41] Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens, A.-R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz, *PLoS Biol.* **7**, e1000049 (2009).
- [42] A. M. de Lachapelle and S. Bergmann, *Mol. Syst. Biol.* **6**, 351 (2010).
- [43] L. Wolpert, *J. Theor. Biol.* **25**, 1 (1969).
- [44] L. Wolpert, *Dev. Genet.* **15**, 485 (1994).
- [45] L. Wolpert, *Trends Genet.* **12**, 359 (1996).
- [46] L. Wolpert, *J. Theor. Biol.* **269**, 359 (2011).
- [47] K. Hironaka and Y. Morishita, *Curr. Opin. Genet. Dev.* **22**, 553 (2012).
- [48] G. Tkačik, J. O. Dubuis, M. D. Petkova, and T. Gregor, *Genetics* **199**, 39 (2015).
- [49] H. Hasegawa, *J. Phys. Soc. Jpn.* **75**, 033001 (2006).
- [50] H. Hasegawa, *Physica A* **374**, 585 (2007).
- [51] R. Rodriguez and H. C. Tuckwell, *Phys. Rev. E* **54**, 5585 (1996).
- [52] D. T. Gillespie, *J. Chem. Phys.* **113**, 297 (2000).
- [53] J. Hattne, D. Fange, and J. Elf, *Bioinformatics* **21**, 2923 (2005).
- [54] J. Elf and M. Ehrenberg, *Syst. Biol.* **1**, 230 (2004).
- [55] G. Tkačik, T. Gregor, and W. Bialek, *PLoS ONE* **3**, e2774 (2008).
- [56] G. Tkačik and W. Bialek, *Phys. Rev. E* **79**, 051901 (2009).
- [57] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [58] A. N. Kolmogorov, *Probl. Pered. Inf.* **1**, 3 (1965).
- [59] T. Bollenbach, P. Pantazis, A. Kicheva, C. Bökel, M. González-Gaitán, and F. Jülicher, *Development* **135**, 1137 (2008).
- [60] A. Kicheva, P. Pantazis, T. Bollenbach, Y. Kalaidzidis, T. Bittig, F. Jülicher, and M. González-Gaitán, *Science* **315**, 521 (2007).

- [61] D. M. Holloway, F. J. P. Lopes, L. da Fontoura Costa, B. A. N. Travenolo, N. Golyandina, K. Usevich, and A. V. Spirov, *PLoS Comput. Biol.* **7**, e1001069 (2011).
- [62] D. M. Holloway and A. V. Spirov, *PLoS ONE* **10**, e0118450 (2015).
- [63] T. Taillefumier and N. S. Wingreen, *PLoS Comput. Biol.* **11**, e1004238 (2015).
- [64] M. Tikhonov, S. C. Little, and T. Gregor, [arXiv:1501.07342](https://arxiv.org/abs/1501.07342).
- [65] D. Cottrell, P. S. Swain, and P. F. Tupper, *Proc. Natl. Acad. Sci. USA* **109**, 9699 (2012).
- [66] K. Kaizu, W. de Ronde, J. Paijmans, K. Takahashi, F. Tostevin, and P. R. ten Wolde, *Biophys. J.* **106**, 976 (2014).
- [67] J. Paijmans and P. R. ten Wolde, *Phys. Rev. E* **90**, 032708 (2014).
- [68] H. C. Berg and E. M. Purcell, *Biophys. J.* **20**, 193 (1977).
- [69] W. Bialek and S. Setayeshgar, *Proc. Natl. Acad. Sci. USA* **102**, 10040 (2005).
- [70] G. Tkačik and W. Bialek, [arXiv:1412.8752](https://arxiv.org/abs/1412.8752).
- [71] H. B. Barlow, in *Sensory Communication*, edited by W. A. Rosenbluth (MIT Press, Cambridge, 1961), pp. 217–234.
- [72] D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek, *Proc. Natl. Acad. Sci. USA* **111**, 3683 (2014).
- [73] A. S. Rajkumar, N. Dénervaud, and S. J. Maerkl, *Nat. Genet.* **45**, 1207 (2013).