Extending the dynamic range of transcription factor action by translational regulation

Thomas R. Sokolowski, Aleksandra M. Walczak, William Bialek, and Gašper Tkačik

Extending the dynamic range of transcription factor action by translational regulation

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

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

A crucial step in the regulation of gene expression is binding of transcription factor (TF) proteins to regulatory sites along the DNA. But transcription factors act at nanomolar concentrations, and noise due to random arrival of these molecules at their binding sites can severely limit the precision of regulation. Recent work on the optimization of information flow through regulatory networks indicates that the lower end of the dynamic range of concentrations is simply inaccessible, overwhelmed by the impact of this noise. Motivated by the behavior of homeodomain proteins, such as the maternal morphogen Bicoid in the fruit fly embryo, we suggest a scheme in which transcription factors also act as indirect translational regulators, binding to the mRNA of other regulatory proteins. Intuitively, each mRNA molecule acts as an independent sensor of the input concentration, and averaging over these multiple sensors reduces the noise. We analyze information flow through this new scheme and identify conditions under which it outperforms direct transcriptional regulation. Our results suggest that the dual role of homeodomain proteins is not just a historical accident, but a solution to a crucial physics problem in the regulation of gene expression.

I. INTRODUCTION

Cells control the concentration of proteins in part by regulating transcription, the process by which mRNA molecules are synthesized from the DNA template. Central to the regulation of transcription are the "transcription factor" (TF) proteins that bind to specific sites along the DNA and enhance or repress the expression of nearby genes. Perhaps surprisingly, many TF molecules are present at very low concentration, and even at low total copy number [1]. While it has been appreciated for many years that low concentrations of biological signaling molecules must lead to significant noise levels [2], direct measurements of the fluctuations in gene expression have become possible only in the past fifteen years [3].

Pathways for the regulation of gene expression can be seen as input-output devices, with information flowing from input control signals (TF concentrations) to output behaviors (number of synthesized protein molecules). While "information" usually is used colloquially in describing biological systems, the mutual information between input and output provides a unique, quantitative measure of the performance of these systems [4, 5]. In the context of embryonic development, for example, the information (in bits) carried by gene expression levels sets a limit on the complexity and reproducibility of the body plans that can be encoded by these genes [6].

Decades of work on neural coding provide a model for the use of information theory in exploring signaling processes in biological systems [6, 7]. To exploit this concept, as a first step it is necessary to estimate the various information theoretic quantities from data on real

systems, and for genetic regulatory networks that has been achieved only very recently. There are estimates of the mutual information between the concentration of a TF and its target gene expression [8, 9], the information that expression levels of multiple genes carry about the position of cells in the developing fruit fly embryo [10, 11], and the information that gene expression levels provide about external signals in mammalian cells [12, 13]. As a second step, we need to understand theoretically how the various features of the systems—the architecture of signal transmission, the noise levels, the distribution of input signals—contribute to determining information transmission. In qualitative terms, the noise levels set a limit to information flow given a fixed maximum signal level, and thus understanding information transmission is intimately connected to the question of how the cell can maximize the information conveyed by a limited number of molecules produced and transported stochastically [14–24]; completing the circle, this problem is directly analogous to the "efficient coding" problem in neural systems [25]. As emphasized in Ref. [5], information theoretic ideas can thus be used as tools for the quantitative characterization of biological systems, but there is also the more ambitious goal of building a theory in which the behavior of real neural, genetic, or biochemical networks could be derived, quantitatively, from the optimization of information flow.

Transmitting maximum information with a limited number of molecules requires regulatory networks to embody strategies for minimizing the effects of noise. Importantly, there are (at least) two contributions to the noise [26], and optimal networks find a balance between these. In transcriptional regulation, the most common gene regulatory mechanism, a commonly appreciated component of noise comes from the stochastic birth and death of the synthesized protein and mRNA molecules [27], which we refer to as "output noise." But there is

---

also noise at the input of the regulatory process, from the random arrival of TF molecules at their binding sites. We can think of transcriptional regulation as a mechanism for sensing the concentration of TFs, connecting the analysis of "input noise" to the broader problem of limits on biochemical signaling and sensing [28–37], first studied in the context of bacterial chemotaxis [2]. Changing the shape of input-output relations, both through cooperativity and through feedback, changes the balance between input and output noise, thus rendering the optimization of information flow a well–posed problem even with very simple physical constraints on the total mean number of molecules [16–18].

Central to any account of noise reduction is the effect of averaging. Several averaging strategies that make transcriptional regulation more reliable have been identified: there is averaging over time as molecules accumulate [2, 28, 31–34], averaging over expression levels of multiple genes that are regulated by the same TF [16, 17, 38, 39], and averaging over space as molecules diffuse between neighboring cells or nuclei, e.g., in a developing embryo [19, 40–43] or organoid [44]. In the (typical) case where one TF targets multiple genes, there is a regime where information transmission is optimized by complete redundancy in the response of these targets, and another regime in which the concentrations for activation or repression of the targets are staggered so as to "tile" the dynamic range of inputs [16, 17]. But, even as we consider networks with increasing numbers of targets, the theoretically optimal strategy is to insert the additional genes into the high-concentration end of the input range and to avoid the lower part of the dynamic range altogether. It is the high input noise at low concentrations that renders this regime suboptimal for reliable signaling.

Here we explore a distinct averaging strategy which allows transcription factors to access the low end of their input range in a robust manner. Since these regulatory proteins bind to DNA, it is plausible that they could also bind to mRNA, thereby regulating translation; this is known to happen in the large class of homeodomain proteins [45–47] and for the Argonaute family proteins [48, 49], for several other proteins that fulfill important functions in the *Drosophila* embryo and oocyte [50–57], but also in other eukaryotic species [58–60] and in prokaryotes [61–67]. Intuitively, each mRNA molecule could act as an independent sensor of the input concentration, and averaging over these multiple sensors could reduce the input noise and thereby allow for more effective information transmission at low input concentrations.

To develop this intuition, we first consider a model of the "direct transcriptional regulation" (DTR) scheme, in which a TF concentration is read out and averaged by M binding sites on the same promoter (Section II); we subsequently generalize the model to a more complicated "indirect translational regulation" (ITR) scheme, in which the averaging function is served by M cytoplasmic mRNA molecules (Section III). We compare the two

regulation mechanisms, DTR and ITR, by computing the maximum information flow in each as a function of the input noise magnitude and other determinants of the information flow (Section IV). We conclude by discussing a biologically relevant example from early *Drosophila* development (Section V).

II. AVERAGING OVER NEIGHBORING REGULATORY REGIONS IN DIRECT TRANSCRIPTIONAL REGULATION

The intuition behind the arguments of this work is that a cell can extract more information from low concentrations of TFs by averaging over multiple binding regions. We expect that this will be realized by having the multiple binding regions on different mRNA molecules. As a motivating exercise, however, we can imagine that there are many regions for binding of the TF at a single target near the gene being regulated, and that the expression of this gene depends on the average of the occupancies of these regions (see schematic in Fig. 1A); there are hints that such non–cooperative regulation by a cluster of binding regions may be realized in some cases [68]. We expect that, with averaging over M binding regions, we should find a $\sqrt{M}$ reduction in noise levels, and our goal here is to exhibit this explicitly, as well as to understand the conditions for this reduction to be achieved. These results will provide a guide to the more complex case of "indirect translational regulation" (ITR), introduced in Sec. III. The calculational framework we use here is based on our previous work [4, 16–18].

We write the expression level of the single target gene as g, and if expression is controlled by the average of multiple nearby regulatory regions then the dynamics are of the form

$$\frac{dg}{dt} = r\left[\frac{1}{M}\sum_{i=1}^{M} f_i(c)\right] - \frac{1}{\tau}g + \xi \;, \qquad (1)$$

where r is the maximal rate of synthesis, $1/\tau$ is the rate at which the gene products are degraded, and ξ is a Langevin noise source (zero-mean white noise). In this model there is a single TF species, at concentration c, that controls expression. We assume τ to be the longest time scale in the problem, thus setting the averaging time for all noise sources in the regulatory pathway. As described more fully in Refs. [8, 15–19], we can think of the regulatory mechanism as propagating information from c to g, and this information transmission is a measure of the control power achieved by the system.

In the simplest case, each region harbors just one binding site, and the contribution of that site to the activation of gene expression is determined by its equilibrium occupancy $\bar{n}_i \in [0, 1]$; then we have

$$f_i(c) = \bar{n}_i(c) = \frac{c}{c + K_i} \;, \qquad (2)$$

where K_i is the binding constant or affinity of site i for the TF. Alternatively, each region, corresponding to a regulatory sequence along the DNA, could be a tight cluster of binding sites that act cooperatively, so that

$$f_i(c) = \frac{c^{H_i}}{c^{H_i} + K_i^{H_i}} \ , \tag{3}$$

where H_i is the Hill coefficient describing the cooperativity. Note that in this parameterization, we can describe activators and repressors by the same equation, using positive and negative H_i, respectively.

The noise term ξ should include many different microscopic effects. There is noise in the synthesis and degradation of the gene product (output noise), and there is noise in the arrival of the TF at its target site (input noise). As in Refs. [15–19], we describe the output noise as a birth–death process, and subsume several complexities by assuming that we are counting "independent events" without making a detailed commitment about their nature (e.g., whether the mRNA or protein molecules are independent, or if the truly independent events are bursts of transcription [69]); we will, however, consider these aspects in more detail in the ITR model (Sec. III).

For the input noise, there is a minimum level set by the Berg–Purcell limit [2, 28, 34, 37], which is equivalent to a variance in the TF concentration,

$$\sigma^2_{c,(\text{in})} = \frac{c}{D_c \ell_c \tau} \Phi_B(\vec{n}(c)) \ , \tag{4}$$

where D_c is the diffusion constant of the TF, ℓ_c is the effective linear dimension of the binding region, τ is the integration time over which noise is averaged, and the "occupancy factor" $\Phi_B(\vec{n}(c))$ is a function of the average occupancy that depends on molecular details, with B denoting the number of binding sites per regulatory region. In particular, for a single binding site with equilibrium occupancy $\bar{n}_1$, we expect $\Phi_1 = (1-\bar{n}_1)^{-1}$ [34, 35]; for a cluster of sites in the limit as their number grows large, such that a single site is never fully saturated, $\Phi_B \to \Phi_\infty \equiv 1$ [28, 70].

We can cast both input and output noise into the Langevin form (cf. Eq. 1), but we know from Refs. [15–19] that, so long as they are not too large, these noise sources provide additive contributions to the variance of g. We can find the effect of the input noise by "propagating errors" through the mean input-output relations, and then add to the output noise:

$$\sigma^2_g = \bar{g} + \sum_{i=1}^{M} \left(\frac{\partial \bar{g}}{\partial f_i}\right)^2 \left(\frac{\partial f_i}{\partial c}\right)^2 \sigma^2_{c,(\text{in})} \ . \tag{5}$$

The first term is the Poisson output noise, with the variance equal to the stationary mean, which can be computed from Eq. (1):

$$\bar{g}(c) = \frac{r\tau}{M} \sum_{i=1}^{M} f_i(c) \ . \tag{6}$$

A Direct transcriptional regulation (DTR)

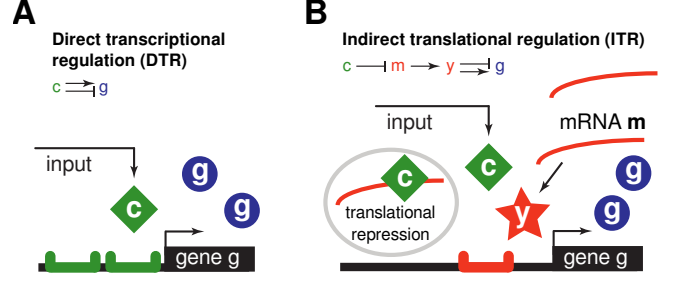**B** Indirect translational regulation (ITR)

FIG. 1. (Color online) **Schematic comparison of direct transcriptional regulation (DTR) and indirect translational regulation (ITR) schemes. (A)** In direct transcriptional regulation (DTR), activator (or repressor) TFs, depicted as green squares and present at concentration c, interact with (potentially multiple, not necessarily identical) TF binding regions to activate (repress) expression of the regulated gene g. **(B)** In the indirect translational regulation (ITR) scenario, input molecules (green squares) bind to mRNAs m of protein y (red chain) to make the mRNA unaccessible for translation (gray oval). Translation can proceed from unbound mRNA molecules, giving rise to proteins y (red stars). These proteins act as repressors (or activators) for gene g; the overall mapping from c to g is thus activating (repressing) in both scenarios.

Note that since $f_i \in [0, 1]$, the maximum mean number of output molecules is $N_{\max} = r\tau$.

Let us now assume, for simplicity, that all regulation functions f_i are identical: all $H_i = H$, all $K_i = K$, and hence all $f_i(c) \equiv f(c) = \bar{n}(c)$. The total noise in gene expression in a model with M identical binding regions then reads:

$$\sigma^2_g = \bar{g} + M \left(\frac{r\tau}{M}\right)^2 \left(\frac{\partial f}{\partial c}\right)^2 \frac{c}{D_c \ell_c \tau} \Phi_B \ . \tag{7}$$

Introducing, consistently with our previous work, a dimensionless concentration unit $c_0 = N_{\max}/D_c\ell_c\tau$, and measuring expression levels g in units of maximal induction $N_{\max} = r\tau$, we observe that the mean expression is simply $\bar{g} = f(c)$, and the noise can be written as

$$\sigma^2_g = \frac{1}{N_{\max}} \left[\bar{g} + \frac{\Phi_B}{M} c \left(\frac{\partial \bar{g}}{\partial c}\right)^2\right] \ . \tag{8}$$

If the input concentration c has a limited dynamic range, i.e., $c \in [0, C]$, where $C = c_{\max}/c_0$ is the maximal allowed concentration of the input in units of c_0, the relative importance of the two noise terms is set by C. For $C \gg 1$, it is possible to regulate the gene such that the input noise contribution [second term of Eq. (8)] is negligible compared to the output noise [first term of Eq. (8)]. For $C \ll 1$, the input noise is dominant and the output noise is negligible, unless M is large. The balancing of these noise terms has been explored in detail in our previous work [15–19].

Alternatively, the total noise at the output from Eq. (8) can be mapped to an equivalent noise at the input

through the slope of the input-output relation $\bar{g}'(c)$,

$$\sigma_c^2 = \frac{1}{N_{\max}} \left[\bar{g} \left(\frac{\partial \bar{g}}{\partial c} \right)^{-2} + \frac{\Phi_B}{M} c \right] \ . \tag{9}$$

Equations (8,9) contain two differences compared to the single input/single output case reported in Ref. [16]. First, the "occupancy factor" Φ_B is introduced by a refinement of the expression for the input noise [34, 35]; this change will not qualitatively influence our conclusions. Moreover, as mentioned above, the significance of this correction decreases as the number of binding sites per region increases [34]. Second, the factor $1/M$ multiplying the input noise contribution suggests that the input noise can be decreased by averaging over multiple binding regions for the TF.

If the transcriptional regulatory apparatus really were driven to strongly suppress the input noise, e.g., by a factor of order 10, this would necessitate large M, and it is hard to imagine how $10^2 - 10^3$ binding regions could be packed into a linear regulatory section on the DNA. One difficulty with this is that there is no plausible molecular machinery which could read out the average occupancy of so many regions. The other, more fundamental difficulty is that, due to close packing on the DNA, such regulatory regions would interact and thus fail to provide independent concentration measurements, likely negating the apparent benefits of input noise averaging. This effect is well known since the original work of Berg and Purcell in the context of chemoreception [2].

In the following section we will show that translational regulation implements an input noise reduction mechanism that is conceptually identical to that of M binding regions, while automatically removing the two associated problems discussed above. We will compare the noise reduction in the "indirect translational regulation" (ITR) mechanism to the "direct transcriptional regulation" (DTR) case, which we define as the simple scheme using $M = 1$ regulatory element with noise given by Eqs. (8,9).

III. INDIRECT TRANSLATIONAL REGULATION

In the indirect translational regulation (ITR) scenario, the gene g is not regulated directly, but through an intermediate step. Let us assume that regulatory protein c translationally represses the mRNA m of a protein whose copy number we will denote by y; this protein acts as a repressing (or activating) TF for the output protein g, as depicted in Fig. 1B. As a result, the end transformation of inputs c to outputs g is again activating (or repressing), and can be compared to the respective direct transcriptional regulation pathway (see Fig. 1).

One possible reaction scheme for indirect translational regulation consists of the following system of equations:

$$\frac{dm}{dt} = r_m - \frac{1}{\tau_m} m - k_+ cm + k_- b \tag{10}$$

$$\frac{db}{dt} = k_+ c\, m - k_- b - \frac{1}{\tau_m} b \tag{11}$$

$$\frac{dy}{dt} = r_y m - \frac{1}{\tau_y} y \tag{12}$$

$$\frac{dg}{dt} = r f(y/\Omega) - \frac{1}{\tau} g \ . \tag{13}$$

Here, the mRNA of the intermediary gene y is produced at rate r_m and degraded at rate τ_m^{-1}. It can be bound by the input translational regulator at rate $k_+ c$, and unbound at rate k_-. The variable m tracks the number of unbound mRNAs from which translation can proceed; b tracks the repressed (bound) mRNA number. Translation of the unbound mRNA occurs at rate r_y and the protein is degraded with rate τ_y^{-1}. Finally, y controls the expression of g, as the input c does in the DTR scenario, through a regulatory function f: g proteins are expressed with maximal rate r, and degraded with rate $1/\tau$, by assumption the slowest time scale in the problem. Here m, b and y are given as absolute molecule counts, c is given in concentration units. The function $f(\cdot)$ is defined to take concentration as input to parallel the DTR scenario; the expression level of the intermediary gene y therefore must be divided by the relevant reaction volume, Ω, when inserted into f. Equations (10-13) still have to be supplemented by the associated Langevin noise terms; we analyze the noise in detail below.

By solving Eqs. (10-13) in the steady state, we obtain the average levels of signaling molecules:

$$\bar{m} = \frac{r_m \tau_m}{1 + \frac{c}{K_c}} \ ,$$
$$\bar{y} = r_y \tau_y \bar{m} \ ,$$
$$\bar{g} = r\tau f\left(\frac{\bar{y}}{\Omega} \right) \ . \tag{14}$$

Here, $K_c = \frac{1 + k_- \tau_m}{k_+ \tau_m}$, and in the limit of fast binding and unbinding ($k_- \tau_m \gg 1$) this reduces to $K_c \approx k_-/k_+$, which is akin to the familiar form for the dissociation constant. As before, we define $N_{\max} = r\tau$ to be the maximum number of molecules at the output. Analogously, let $M = r_m \tau_m$ be the steady state number of mRNA, either active (unbound by c) or repressed (bound by c), i.e., $M = \bar{m} + \bar{b}$.

To compute the noise in gene expression at steady state, we consider the following noise sources: (i) shot noise due to the production and degradation of g; (ii) shot noise due to the production and degradation of the protein y; (iii) shot noise due to production and degradation of the mRNA m; (iv) diffusion noise due to random arrival of y molecules at the promoter of g; and (v), diffusion noise due to random arrival of c molecules at the mRNA of y. Only two of these noise sources [(i) and (v)] arise from directly analogous processes in the DTR

scheme. The additional sources reflect the increased complexity of the ITR scheme, suggesting that ITR will only be beneficial over DTR when the reduction in the input noise of c (v) is large enough to compensate for the effect of the new noise sources.

As in previous work (Refs [15–19], but see also [20]), this analysis neglects switching noise, i.e., the noise due to stochastic transitions between different occupancy states of the promoter [28, 71–75]. One reason for this is that such noise contributions depend on the exact molecular mechanisms operational at the promoter, which we do not know in detail. The other reason is that the noise contributions analyzed here represent the physical bounds due to finite number of molecules. At sufficiently low concentrations, which are the focus of interest here, these noise contributions will overwhelm the switching noise; more generally we can imagine that cells have evolved mechanisms that minimize these adjustable sources of noise, leaving the physically inevitable noise sources to dominate. We will now analyze the effect of the different noise contributions on the steady state variance of the output g in detail.

Birth-death noise sources (i)–(iii). To correctly compute the birth-death (shot) noise contributions to the total noise at the output, it is instructive to consider the propagation of arbitrary shot noise sources in a generic signaling cascade of the form:

$$\frac{dx_1}{dt} = r_1 - \frac{1}{\tau_1}x_1 + \xi_1$$
$$\frac{dx_2}{dt} = r_2 f_2(x_1) - \frac{1}{\tau_2}x_2 + \xi_2$$
$$\vdots$$
$$\frac{dx_n}{dt} = r_n f_n(x_{n-1}) - \frac{1}{\tau_n}x_n + \xi_n \ , \qquad (15)$$

where the shot noise spectra are

$$\langle \xi_j(t)\xi_k(t')\rangle = 2\tau_j^{-1}\bar{x}_j\delta_{jk}\delta(t-t') \qquad (16)$$

in the steady state, with $\bar{x}_j = \langle x_j\rangle$ the stationary mean. Assuming that $\tau_n \gg \tau_j$ for all $j \neq n$, it can be shown by linearizing and Fourier transforming Eqs. (15) (see Appendix A) that the total variance in x_n, to a good approximation, is $\sigma_n^2 = \sum_{j=1}^{n-1}\sigma_{n\leftarrow j}^2 + \bar{x}_n$, where

$$\sigma_{n\leftarrow j}^2 = \left[\prod_{q=j+1}^{n} r_q\tau_q f_q'\right]^2 \times \left(\frac{\tau_j}{\tau_n}\right) \times \bar{x}_j \ . \qquad (17)$$

Equation (17) is intuitive to interpret: shot noise entering the cascade at step j has variance equal to the mean, $\bar{x}_j$, which gets filtered by the temporal averaging, τ_j/τ_n, over what is ultimately the slowest timescale in the problem, $\tau_n = \tau$, and is finally propagated through all subsequent

stages of the signaling pathway, given by the gain factors and slopes of the input-output relations [76].

Diffusion noise sources (iv) and (v). Here we first note that the contribution of diffusion noise sources to the variance in the output can be generically written as

$$\sigma_{g\leftarrow x}^2 = \left(\frac{\partial\bar{g}}{\partial x}\right)^2 \frac{x}{D_x\ell_x\tau}\Phi_x \ , \qquad (18)$$

where x is the concentration of the diffusing species, Φ_x is a function of the internal state of the regulatory region to which x is binding, ℓ_x is its linear extent, and the derivative acts on the steady-state transformation between x and the mean output $\bar{g}$, given by Eqs. (14).

For noise source (iv), the diffusive species is y, and the target is the TF binding site controlling the expression of g. The relevant input-output relation through which the noise is propagated is $\frac{\partial\bar{g}}{\partial(y/\Omega)} = r\tau f'(y/\Omega)$. This yields:

$$\sigma_{g\leftarrow y}^2 = r^2\tau^2\left[f'(\bar{y}/\Omega)\right]^2 \frac{(\bar{y}/\Omega)\ \Phi_y}{D_y\ell_y\tau} \ . \qquad (19)$$

To allow for flexible regulation that can implement different input-output curves $f(y/\Omega)$, we will later assume that the expression is effected by many binding sites, resulting in $\Phi_y \simeq 1$.

For noise source (v), the diffusive species is c, and the targets are mRNAs m. Since each mRNA molecule acts as an autonomous "detector" for the concentration c, the relevant "receptor occupancy" is the probability of a single mRNA to be bound and thus repressed by c; denoting the average single-mRNA occupancy as $\bar{m}_1 \equiv \bar{m}/M$, we can write the diffusion noise for each single mRNA as:

$$\sigma_{m_1}^2 = \left(\frac{\partial\bar{m}_1}{\partial c}\right)^2 \frac{c\ \Phi_c}{D_{cm}l_c\tau} = \left(\frac{1}{M}\frac{\partial\bar{m}}{\partial c}\right)^2 \frac{c\ \Phi_c}{D_{cm}l_c\tau} \ , \quad (20)$$

where $D_{cm} = D_c + D_m$, accounts for the fact that in the ITR scenario also the detectors can be mobile; usually we can expect $D_c \gg D_m$, such that $D_{cm} \simeq D_c$, and in the following we thus set $D_{cm} = D_c$. As in the direct regulation model, we can add up the diffusion noise for the identical but independent c-detectors to obtain the diffusion noise in the total mRNA population: $\sigma_m^2 = M\sigma_{m_1}^2$. Propagating this noise through the downstream input-output relations finally yields the expression for noise source (v):

$$\sigma_{g\leftarrow c}^2 = \left(\frac{\partial\bar{g}}{\partial\bar{m}}\right)^2 M\left(\frac{1}{M}\frac{\partial\bar{m}}{\partial c}\right)^2 \frac{c\ \Phi_c}{D_cl_c\tau} \ . \qquad (21)$$

Assuming a single TF binding site per mRNA, we can expect $\Phi_c = (1 - \bar{b}/M)^{-1}$ [34], which we use in our subsequent calculations.

Assembling all noise sources together. Applying the above considerations to the ITR regulation sheme defined by Eqs. (10-13), we can write down the steady state variance in the output as:

$$\sigma_g^2 = \underbrace{\bar{g}}_{(i)} + \underbrace{\left(\frac{rf'}{\Omega}\right)^2 \tau\tau_y\bar{y}}_{(ii)} + \underbrace{\left(\frac{rf'\tau}{\Omega}\right)^2 (r_y\tau_y)^2 \left(\frac{\tau_m}{\tau}\right)\bar{m}}_{(iii)} + \underbrace{\frac{\bar{y}}{D_y\ell_y\tau\Omega}\Phi_y r^2\tau^2 f'^2}_{(iv)} + \underbrace{\frac{c}{D\ell_c\tau}M\Phi_c\left(\frac{1}{M}\frac{\partial\bar{m}}{\partial c}\right)^2\left(\frac{\partial\bar{g}}{\partial\bar{m}}\right)^2}_{(v)} \ . \quad (22)$$

Let us now choose a set of units that is natural for the ITR scheme and consistent with the DTR scenario. As before, we measure the output $\bar{g}$ in units of $N_{\max} = r\tau$ such that it falls into $[0,1]$, and we measure the concentration c in units of c_0,

$$c_0 = \frac{N_{\max}}{D_c\ell_c\tau} \ . \quad (23)$$

In direct analogy, we choose a concentration unit for proteins y:

$$y_0 = \frac{N_{\max}}{D_y\ell_y\tau} \ . \quad (24)$$

Since the binding sites and diffusion constants for TFs y and c are in principle different, these units could be different, but we will later assume them to be similar. Let us also define:

$$y_{\max} = \frac{r_y\tau_y}{\Omega y_0} \quad (25)$$

as the maximum (dimensionless) concentration of proteins y expressed from a *single* mRNA.

Binding and unbinding of c to mRNA defines the first nonlinearity of the problem; denoting the average occupancy of a *single* mRNA molecule by $h(c)$, we have

$$h(c) = \frac{K_c}{K_c + c} \ , \quad (26)$$

where c and $K_c = (1 + k_-\tau_m)/(k_+\tau_m c_0)$ are both dimensionless. Note that $h(c) = \bar{m}/M$. The second nonlinear function is $f(y)$, determining the output expression level of g (as in the DTR scenario); its argument is a dimensionless concentration of y, measured in units of y_0. We will explore the space of functions with a Hill form

$$f(y) = \frac{K^H}{y^H + K^H} \ , \quad (27)$$

where K is also measured in units of y_0, and the Hill coefficient H roughly corresponds to the number of y binding sites on the promoter of g. Here again the sign of H determines whether y is an activator or repressor to g. In this work, we will focus on the case $H < 0$ (meaning that y activates g, while itself being repressed by the upstream input c, resulting in *overall* negative regulation of g by c), and compare to the DTR model in which g is repressed by c as well [$H < 0$ in Eq. (3)].

We can rewrite the noise in Eq. (22) as the effective noise at the input,

$$\sigma_g^2 = \left(\frac{\partial f}{\partial(\bar{y}/\Omega y_0)}\right)^2\left(\frac{\partial(\bar{y}/\Omega y_0)}{\partial\bar{m}}\right)^2\left(\frac{\partial\bar{m}}{\partial c}\right)^2\sigma_c^2 = y_{\max}^2 M^2 f'^2 h'^2 \ \sigma_c^2 \ . \quad (28)$$

After rearranging terms and writing all copy numbers and concentrations in the new units defined above, we obtain

$$\sigma_c^2 = \frac{1}{N_{\max}}\left[\underbrace{\frac{1}{M^2}\frac{f}{f'^2}\frac{1}{h'^2}\frac{1}{y_{\max}^2}}_{(i)} + \underbrace{\frac{F}{M}\frac{h}{h'^2}}_{(ii)+(iii)} + \underbrace{\frac{1}{M}\frac{h}{h'^2}\frac{\Phi_y}{y_{\max}}}_{(iv)} + \underbrace{\frac{\Phi_c}{M}c}_{(v)}\right] \ , \quad (29)$$

where

$$F = \frac{r}{r_y}(1 + r_y\tau_m) \ , \quad (30)$$

and f and its derivative are evaluated at $\bar{y}/\Omega y_0 = y_{\max}\bar{m} = y_{\max}Mh(c)$. As before, the input concentrations can vary across the range $c \in [0, C]$, where $C = c_{\max}/c_0$ is the dimensionless maximal concentration of the input. If gene expression rates for the target and intermediary proteins were similar, $r \sim r_y$, then $F \approx 1 + r_y\tau_m$, which is reminiscent of the Fano factor due to the "burst size" $r_y\tau_m$, the number of proteins of y expressed on average from one mRNA; we will therefore refer to F by "Fano factor" in the following. In a modified variant of the ITR model where mRNAs are not continuously created and degraded to maintain an average of M copies, but are present at a fixed total number of M copies, $F = r/r_y \approx 1$.

It is useful to remind ourselves of the corresponding result in the DTR case, which is Eq. (9) with $M = 1$,

$$\sigma_c^2 = \frac{1}{N_{\max}} \left[\underbrace{\bar{g} \left(\frac{\partial \bar{g}}{\partial c} \right)^{-2}}_{(\mathrm{i}')} + \underbrace{\Phi_B c}_{(\mathrm{v}')} \right] \ . \quad (31)$$

We see that term (i) of the ITR case is equal to term (i′) in the DTR case, although this identification requires us to propagate the inputs through more layers of response in the ITR, hence a more complex expression. Similarly, term (v) of the ITR case, representing diffusive noise for c-molecules binding to mRNA m, is directly analogous to the diffusion noise term (v′) in the DTR case, but ITR reduces the input noise variance by a factor of M. While Eq. (29) appears complicated, we can nevertheless estimate the relative magnitudes of different noise sources and assess their relevance.

Let us first compare the relative magnitude of the two input-type noise sources. The scale of the ITR noise component due to finite number of y molecules (iv) relative to the input noise in c (v) is

$$\frac{(\mathrm{v})}{(\mathrm{iv})} = y_{\max} \frac{h'^2}{h} \frac{\Phi_c}{\Phi_y} c \sim \frac{y_{\max}}{C} \ . \quad (32)$$

We expect that Φ and regulatory functions h are of order unity, and that the natural scale of the concentration c is given by C, the maximal dimensionless input concentration; then the scale of the derivative of h is $h' \sim C^{-1}$, leading to the final result. The importance of this term thus depends on the comparison of the maximal concentration of the intermediary y proteins and the input proteins c. Clearly, if the intermediary proteins of y are present at very low copy numbers, their diffusion noise will become limiting, and the ITR scheme will be ineffectual.

The scale of input noise (v) relative to the components of the ITR noise due to birth-death processes, (ii)+(iii), is given by

$$\frac{(\mathrm{v})}{(\mathrm{ii}) + (\mathrm{iii})} = \frac{h'^2}{h} \Phi_c \frac{c}{F} \sim \frac{1}{FC} \ . \quad (33)$$

As we argue above, we should expect values of $F \geq 1$. The importance of noise sources (ii)+(iii) thus depends on the scale of F relative to C, and in the regime of low input concentration, $C \ll 1$, noise sources (ii)+(iii) will be negligible for models with small transcriptional bursting for y proteins, i.e., when $F \sim 1$.

Finally, we can assess the relative scale of the input noise (v) with respect to the output noise (i):

$$\frac{(\mathrm{v})}{(\mathrm{i})} = M \frac{h'^2 f'^2}{f} y_{\max}^2 \Phi_c c \sim \frac{H^2}{MC} \ . \quad (34)$$

Here, the regulation functions f and h are again of order unity, but the derivative of f has a scale of $H/(My_{\max}h)$,

H being the Hill coefficient of f. As a result, the dependence of the noise term (i) on $y_{\max}$ exactly cancels out; similarly, the M-dependence of term (i) cancels out exactly. This is to be expected: by increasing M, one can average away the input noise, but the magnitude of the output noise can not be reduced. The same is true for Eq. (9), where the output noise contribution is not divided by M. This term is thus important as it must become limiting when M grows large, with the relevant scale being set by $H^2/MC \sim 1$. The quadratic scaling with H shows that steep regulation curves strongly amplify the noise on the input side.

IV. COMPARING OPTIMAL INFORMATION FLOW IN THE TWO SCHEMES

To compare the regulatory power of direct transcriptional regulation (DTR) and indirect translational regulation (ITR), we ask how much information can be transmitted through each scheme. If noise in gene expression is Gaussian, as we assumed here, then the response of the regulatory pathway is fully characterized at steady state by the distribution $P(g|c) = \mathcal{G}(g; \bar{g}(c), \sigma_g(c))$, where $\mathcal{G}$ is a normal distribution with the mean and standard deviation that can be computed from the Langevin equations Eqs. (1,10-13). The mutual information, $I(c; g)$, between the regulator concentration, c, and the downstream expression level, g, is then given by:

$$I(g; c) = \int dc \, P_c(c) \int dg \, P(g|c) \log_2 \left[\frac{P(g|c)}{P_g(g)} \right] \ , \quad (35)$$

where $P_g(g) = \int dc \, P_c(c) P(g|c)$. Mutual information, a non-negative number measured in bits, tells us how precisely the input c determines the output g, given that the noise places limits to the fidelity of this control. This quantity still depends on the distribution of input protein levels, $P_c(c)$, experienced by the regulatory pathway. It is possible to find the optimal distribution $P_c^*(c)$ that maximizes the information, and the maximum achievable information is then referred to as the (channel) capacity [4, 8]. $P_c^*(c)$ is tuned to the noise properties of the regulation process, favoring the use of input concentrations at which the regulatory element responds with smaller noise over those inputs where noise is higher.

While finding the optimal input distribution and the corresponding maximal information I is difficult in general, in the case of small noise, $\sigma_g \ll 1$, we previously derived a simple formula for the capacity [8]:

$$I(c; g) = \log_2 \frac{Z}{\sqrt{2\pi e}} \equiv \log_2 \tilde{Z} + \frac{1}{2} \log_2 N_{\max} \ , \quad (36)$$

where

$$Z = \int_0^C \frac{dc}{\sigma_c(c)} \ , \quad (37)$$

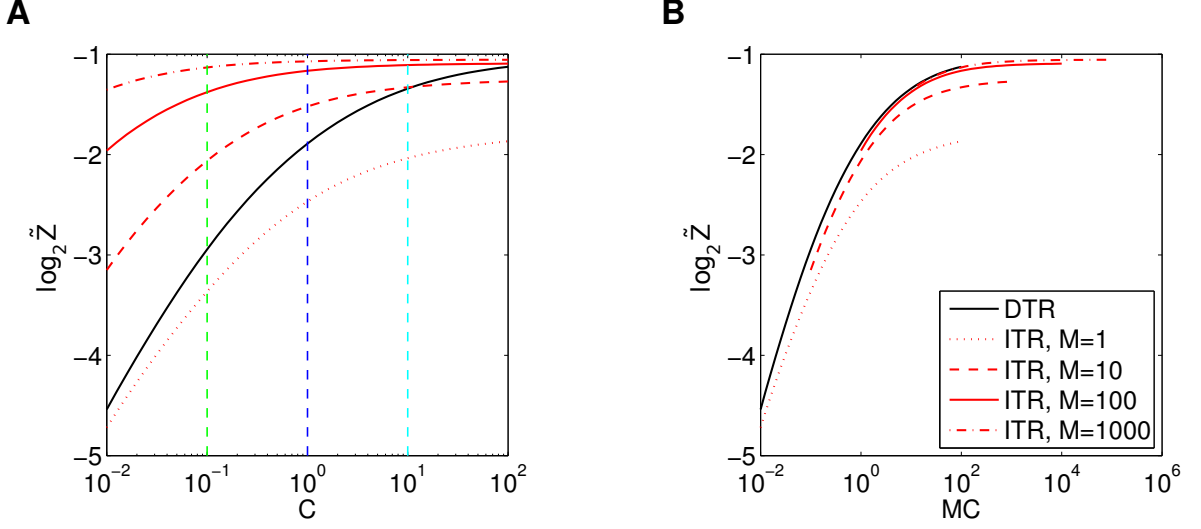$$\tilde{Z} = Z(N_{\max} = 1)/\sqrt{2\pi e} \ , \quad (38)$$

FIG. 2. (Color online) **Optimal capacity of direct transcriptional regulation (DTR, black), compared to indirect translational regulation (ITR, red), as a function of the maximal input concentration C. (A)** Optimal capacity $\log_2 \tilde{Z}$ vs. C for various total amounts of translationally targeted mRNAs, M. Here and in subsequent panels we fix $N_{\max}$, the maximal output copy number, to a reference value $N_{\max} = 1$ and show the resulting capacity; larger values of $N_{\max}$ simply shift all capacities upwards by an additive amount, as in Eq. (36). The strength of "ITR noise" sources, (ii) + (iii) + (iv) of Eq. (29), is set by (Fano factor) $F = 1$ and (maximal concentration of the intermediary protein) $y_{\max} = 10$. For $M > 1$, ITR outperforms DTR at low C, but at high C DTR can still reach higher capacities if M is not sufficiently large (e.g., dashed red line for $M = 10$). At $M = 1$ (dotted red line), the ITR scheme cannot benefit from input noise averaging, yet the intermediary regulatory steps still contribute the "ITR noise" absent in DTR, causing DTR to be superior to ITR at all C. **(B)** Capacity curves for different values of M in (A) collapse when plotted against the product of maximal input concentration and the number of mRNA targets, MC, as predicted by the scaling relation in Eq. (34). Increasing input noise by lowering C thus can be compensated for by increasing M. The collapse is not perfect at high MC and curves for different M saturate at different capacity values because the strength of "ITR noise" is not negligible and their effect on capacity depends on M.

and σ_c is given either by Eq. (9) for the case of direct transcriptional regulation, or by Eq. (29) for the case of indirect translational regulation. Z plays the role of the normalization constant in the distribution over inputs that maximizes $I(c; g)$; the optimal distribution is $P_c^*(c) = Z^{-1}\sigma_c^{-1}(c)$. The simple dependence of capacity on $N_{\max}$ in Eq. (36) follows because $N_{\max}$ only enters the expression as a multiplicative prefactor to σ_c.

The capacity given by Eqs. (36, 37) still depends on the regulatory parameters. We view the maximum concentration of input molecules, C, not as a parameter but as a constraint; in the ITR case there is also a constraint on the number of mRNA molecules, M, and, less importantly, on the statistics of translation, captured by F. In the DTR scheme, the parameters that we can adjust in order to optimize information transmission are the dissociation constant (K) and the Hill coefficient (H) of the regulatory function f in Eq. (1). For ITR, we have the same two parameters determining the properties of the regulatory function f in Eq. (13), plus an extra parameter, K_c, the dissociation constant for the repression function $h(c)$. Since our focus is on the regime where input noise is limiting, $C \ll 1$, we fix $y_{\max}$ and F such that "ITR noise" sources, relating to intermediary com-

pounds, i.e., the mRNA and protein of y, are not dominant. The analysis of noise terms from the preceding section suggests that choosing $F = 1$ and $y_{\max} = 10$ will ensure that noise in the ITR scenario is dominated by the input noise in c [(v)] and the output noise due to production and degradation of g [(i)]. In this regime, the performance of the complex ITR regulatory scheme should approach the direct regulation using M regulatory elements, Eq. (9).

A. Optimal solutions

We mapped the optimal capacity as a function of C for the DTR and ITR scenarios at various values for M, with the results shown in Fig. 2. For each point on the capacity curves, information was maximized with respect to the regulatory parameters ($\{K, H\}$ in case of DTR, $\{K, H, K_c\}$ in the case of ITR). As expected, increasing M clearly enhances the capacity by additionally suppressing dominant input noise at $C < 1$ relative to direct regulation. For $C > 1$, the performance of ITR vs. DTR depends on the choice of $y_{\max}$ and F, which determine when additional noise sources specific to the ITR scheme
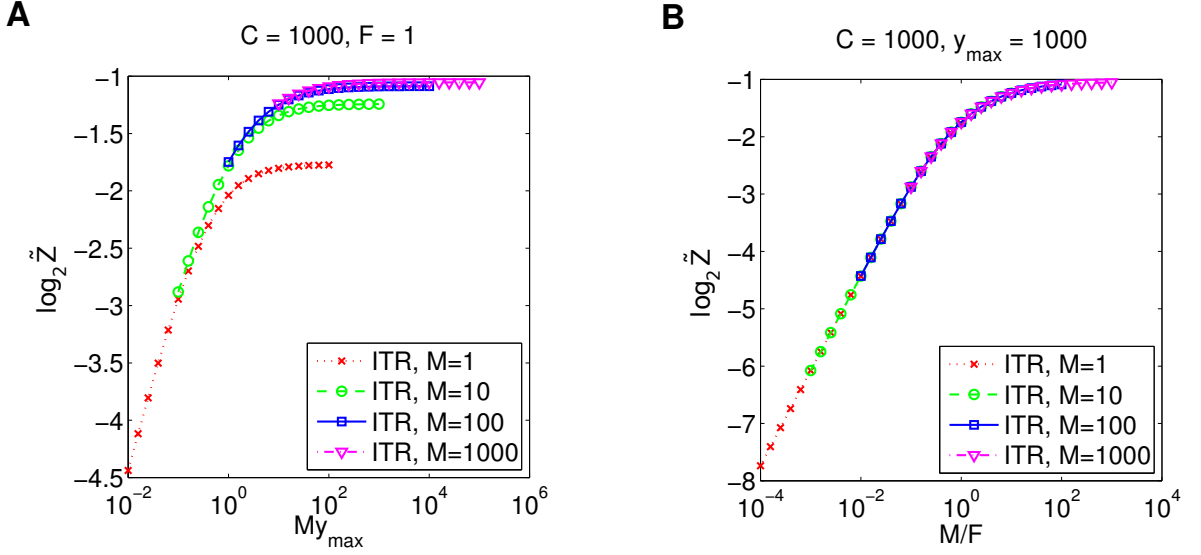
FIG. 3. (Color online) **Scaling of the optimal capacity with "ITR noise" parameters y_{max} and $1/F$.** The optimal capacity $\log_2 \tilde{Z}$ for different fixed values of the mRNA number M in the ITR model is plotted against "ITR noise" parameters properly rescaled by M, to observe the compensation between translational noise components and the input noise. **(A)** For the diffusion noise due to the intermediary protein y, the relevant parameter is $M \times y_{max}$, where y_{max} is the maximal concentration of y. **(B)** For the shot noise due to the expression of intermediary mRNA and protein y, the relevant parameter is $M \times (1/F)$, where $1/F$ is the inverse Fano factor. In both cases, we set the remaining noise sources to be as small as possible: $C = 1000$ and $F = 1$ for (A), $C = 1000$ and $y_{max} = 1000$ for (B). A perfect collapse in (A), comparable to that in (B), could only be achieved for Fano factors $F \ll 1$, which are biologically unrealistic.

become important.

At very small $C \sim 10^{-2}$, the input noise clearly limits capacity in the DTR scenario. As we switch over to the ITR scenario with $M = 10$, we expect the (dominant) input noise variance to drop by a factor of 10, leading to an increase in $I(c; g)$ of $\log_2 \sqrt{10} \approx 1.7$ bits, roughly equal to the observed difference between the DTR curve and the ITR curve for $M = 10$. As M is increased, this scaling breaks down because the reduced input noise stops being the sole factor limiting information transmission, and the capacity curves flatten as a function of C, saturating towards the limit where the transmission is limited only by output and ITR noise. We draw attention to the magnitude of these effects: the capacity of real regulatory elements is in the range of one to several bits [8], so that differences on the order of one bit are huge.

The value of C where the capacity begins to saturate depends on M, because increasing M is equivalent to increasing C as predicted by the scaling relation of Eq. (34). To make this explicit we plotted the same capacity data against MC instead of C, as shown in Fig. 2B. As expected, the curves collapse in the low-MC regime, where input noise is dominant, and start to saturate at similar MC values. However, they do not reach the same plateau at saturation, because the ITR noise contributions, (ii) + (iii) + (iv), are non-negligible. To see that these contributions can be traded off against the decrease in input noise set by M as well, we varied parameters F and y_{max} at fixed $C = 1000$ in Fig. 3, such that only one noise source

remained non-negligible compared to the others. When the resulting capacity is plotted against the appropriately rescaled versions of the parameters, My_{max} and M/F, respectively, we again see a collapse of the capacity curves for different M, as predicted in Section III. In Fig. 3B the collapse is almost perfect, because y-input noise (iv) dominates whereas the respective other two noise sources are negligible, and the plateau thus is purely set by output noise. In Fig. 3A this is not possible because under realistic conditions the Fano factor $F \geq 1$, meaning that the ITR noise cannot be completely eliminated; however, we verified that a perfect collapse can also be achieved here by setting F to unrealistically low values $F \ll 1$.

What do the optimal regulatory curves look like? Figure 4A shows that at low $C = 0.1$, where capacity enhancement by translational regulation is largest, the lowered input noise in the ITR model allows for markedly steeper effective input-output curves $\bar{g}(c)$, especially for $M \gg 1$. At larger C, the difference in capacity between ITR and DTR decreases, as expected from Fig. 2, and the input-output curves converge towards similar effective Hill coefficients in both models, but ITR has a significantly lower midpoint input concentration than DTR. These observations are confirmed by a more systematic analysis, presented in Figs. 4B and C, where we plot the optimal (effective) transition point K_{eff}^* and the optimal Hill coefficient H^*, respectively, as functions of C.
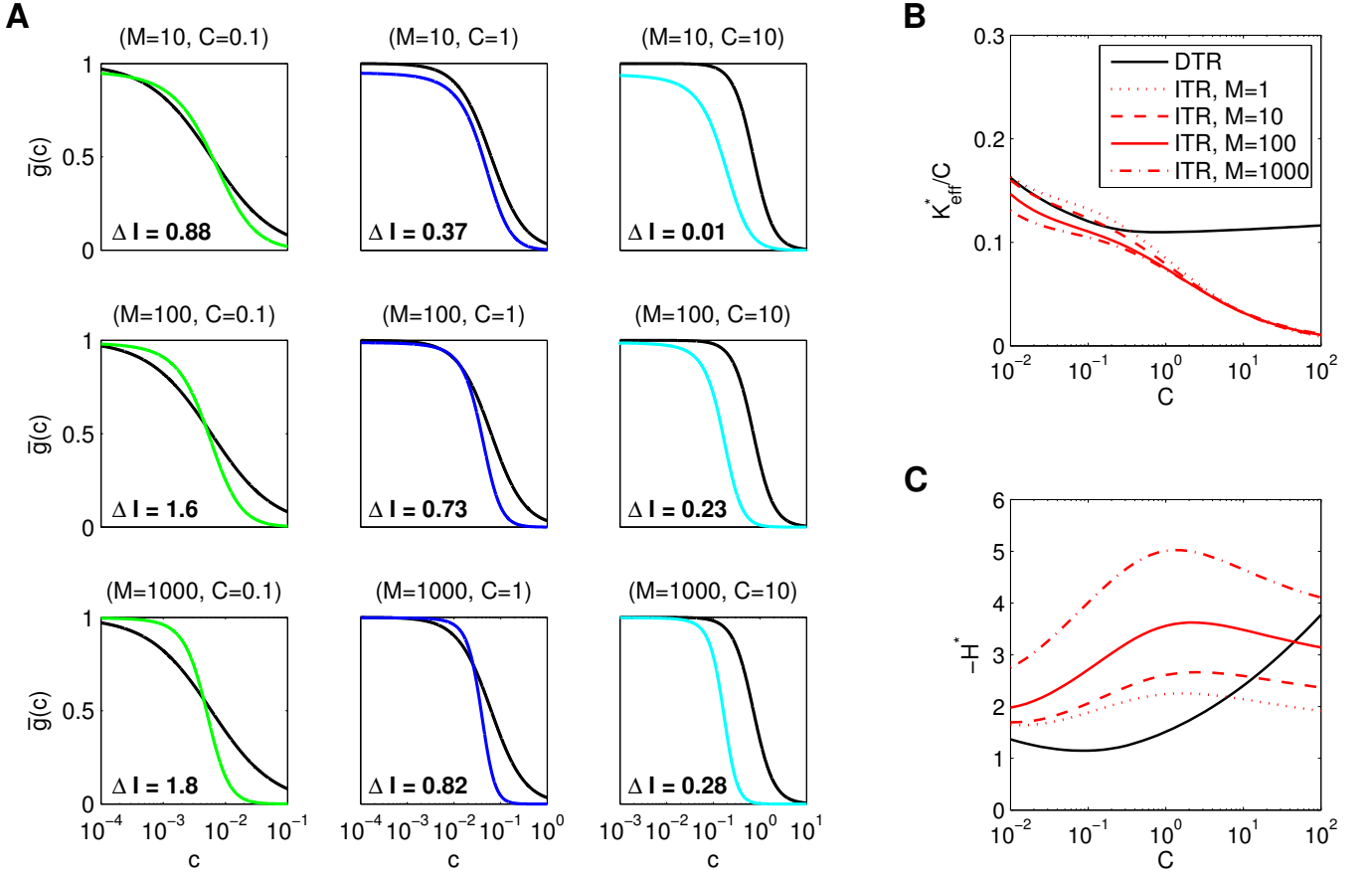
FIG. 4. (Color online) **Optimal regulatory curves.** **(A)** Optimal regulatory curves, $\bar{g}(c)$, for $C = 0.1$ (left), $C = 1$ (middle) and $C = 10$ (right), and different values of M in the indirect translational regulation (ITR) model (colored curves): $M = 10$ (top), $M = 100$ (middle), $M = 1000$ (bottom). The direct transcriptional regulation (DTR) model is plotted in black for reference. Colors correspond to the vertical lines in Fig. 2. Here, $F = 1$ and $y_{\max} = 10$. ΔI, the difference in capacity between ITR and DTR in bits, is shown in each panel. **(B, C)** Optimal regulation threshold, K_{eff}^*, and the optimal Hill coefficient, H^*, that characterize the regulatory curves in (A), shown as a function of C. Since the overall regulation is repressive in either scheme, Hill coefficients are negative. In the DTR model, $K_{\text{eff}} \equiv K$; for ITR, $K_{\text{eff}} = (K_c/K)My_{\max} - K_c$.

B. Comparison at equal resources

In the previous section we compared the optimal operating points for the two regulation scenarios independently, neglecting the resource cost associated with reaching the corresponding optima. However, indirect translational regulation evidently consumes more resources than direct regulation. The extra resource cost is set by M, which determines the cost of the mRNA for the intermediary protein y, as well as $y_{\max}$, which determines the cost of the y protein itself. While it is difficult to convert the cost of mRNA and cost of protein into the same "currency," we can nonetheless compare the costs of the ITR and DTR schemes in terms of the total number of mRNA molecules $\mathcal{M}$ alone.

In the DTR scheme, a good estimate for this cost is $\mathcal{M} = N_{\max}\langle g \rangle$, where $\langle g \rangle = \int g(c)P^*(c)dc$ is the expected number of independent output molecules in response to the optimal distribution of inputs, $P_c^*(c)$. In the ITR scheme, this cost is increased by the number of

mRNA molecules needed to implement translational regulation, M, such that $\mathcal{M} = N_{\max}\langle g \rangle + M$. In this case, $\langle g \rangle$ is implicitly a function of M, because different values of M lead to different noise levels and consequently different optimal input distributions $P^*(c)$.

When we fix the total resource cost $\mathcal{M}$ in both scenarios, the information in the direct scenario is given by:

$$I(c;g) = \log_2 \tilde{Z} + \frac{1}{2}\log_2 N_{\max} = \log_2 \tilde{Z} + \frac{1}{2}\log_2 \frac{\mathcal{M}}{\langle g \rangle} . \tag{39}$$

For translational repression, the (optimized) information reads

$$I(c;g) = \log_2 \tilde{Z}(M^*) + \frac{1}{2}\log_2 \frac{\mathcal{M} - M^*}{\langle g \rangle} , \tag{40}$$

where M^* is the value of $M \in [0, \mathcal{M}]$ that optimizes the capacity, reflecting the optimal allocation of available resources, $\mathcal{M}$, between the translational mechanism that
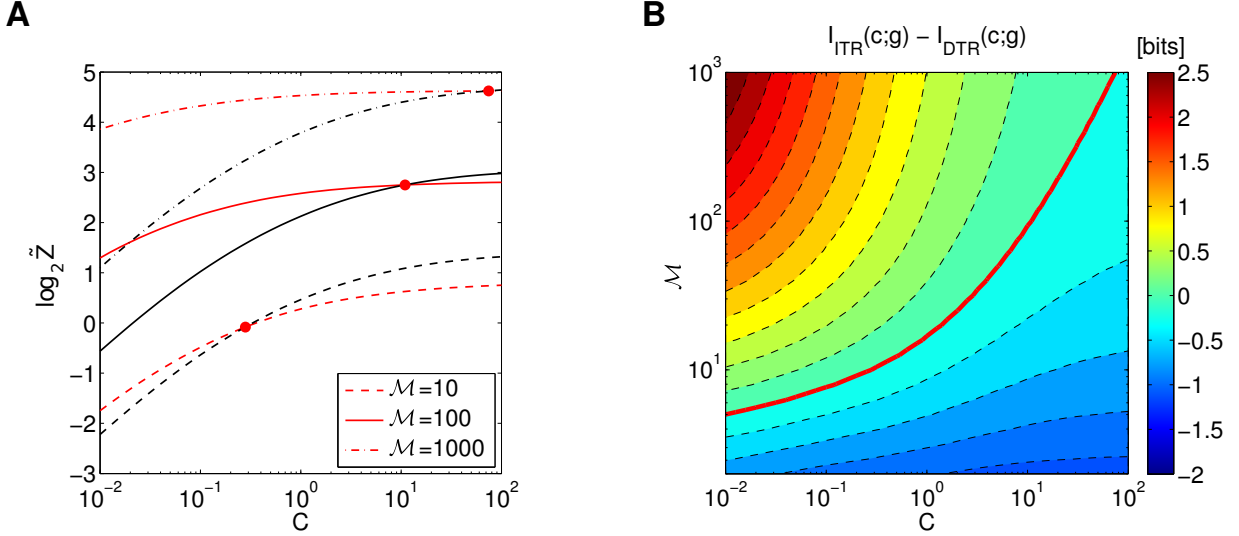
FIG. 5. (Color online) **Direct transcriptional regulation (DTR) vs. indirect translational regulation (ITR) at equal resources. (A)** Comparison of the maximal capacity as a function of C between indirect translational regulation (ITR, red curves) and direct transcriptional regulation (DTR, black curves) scenarios, for three values of total resource cost (total mRNA number), $\mathcal{M}$. Here, $y_{\max} = 10$, $F = 1$. Compared to Fig. 2, ITR stops being beneficial compared to DTR at lower C. **(B)** Capacity difference between the ITR and DTR models as a function of C and $\mathcal{M}$. The thick red line marks the $(C, \mathcal{M})$ values that lead to equal capacities; in the regime above the red line ITR, in spite of intermediary regulatory steps, is beneficial over DTR.

reduces input noise (favoring large M) vs. the reduction of the output noise (favoring large $N_{\max}$, i.e. smaller M). M^* can be found by first optimizing $\log_2 \tilde{Z}(M)$ for fixed M and then optimizing $I(c; g)$ over M in the second step.

Figure 5 compares the information capacities of the ITR (red lines) and DTR (black lines) scenarios for three fixed values of the total mRNA number ("cost") $\mathcal{M}$ $(10, 100, 1000)$. Two effects can be observed: First, in both scenarios the capacities increase markedly with $\mathcal{M}$; this is expected, because increasing $\mathcal{M}$ reduces the overall noise in both scenarios. Note, however, that the gain in capacity upon increasing $\mathcal{M}$ is larger in the ITR scenario (compare red curves) than in the DTR scenario (compare black curves), in particular at low C. This is also expected because the ITR scheme is more efficient in suppressing input noise; in accordance, optimal mRNA values M^* at low C are higher than at high C (data not shown). Second, due to the additional resource requirements of the ITR scheme, it stops being beneficial over the DTR scheme already at lower C, compared to the situation where constrained resources are not taken into account in Fig. 2.

We mapped, systematically, the conditions under which ITR becomes beneficial over DTR. Figure 5B shows the capacity difference $I_{\text{ITR}} - I_{\text{DTR}}$ as a function of the two main factors that influence capacity, C and $\mathcal{M}$. The thick red line marks the combinations $(C, \mathcal{M})$ that lead to equal capacity in both models, i.e. it summarizes the crossing points between ITR and DTR curves in Fig. 5A (thick red points); above the line ITR is superior

to DTR, below DTR yields higher capacities. ITR is beneficial at low C (i.e., high input noise), and as C increases, ITR remains beneficial only for resource-intensive regulatory schemes whose cost $\mathcal{M}$ grows sufficiently quickly with C. We note, once again, the large size of these information differences, in bits.

V. DISCUSSION

An efficient regulatory pathway will respond to variation in its signal across the full input dynamic range. But what if a part of this dynamic range is associated with very high input noise, as is the case for regulatory molecules present at low concentration? Then, the pathway can either avoid responding to those signals completely, thereby sacrificing some of the bandwidth, or it can utilize reaction schemes that are able to reduce the impact of high input noise. In this paper we showed that indirect translational regulation (ITR) is one such regulatory scheme. The intuitive reason for the advantage of translational regulation is that every single one of the M mRNA molecules that is being translationally regulated by the input signal effectively acts as a "receptor" for the input concentration; this results in an M-fold more efficient averaging of input noise compared to the case of direct transcriptional regulation (DTR), where a single DNA binding site is acting alone as a receptor. Consistent with this intuition, our results show that when input noise is dominant ($C \ll 1$), translational repression with

high M can provide large increases in channel capacity relative to direct regulation.

Increases in capacity, however, do not come for free. First, the ITR scheme involves additional reaction steps and intermediary regulatory molecules; as a consequence, new noise sources, which we call "ITR noise," are introduced. Only when these sources are sufficiently small relative to the input noise set by C, does the translational scheme yield measurable benefits. Second, lowering ITR noise sources also incurs metabolic costs associated with producing the required intermediary molecules; this means that a realistic comparison between the direct scheme and the translational repression scheme is relevant only when carried out at comparable resources. We explored these effects in detail to show when translational repression is beneficial over direct regulation.

In this work we have analyzed only one particular reaction scheme for translational regulation, but clearly many variations on the same idea are possible. We examined several of these possibilities, and found consistent results. First, the transformation of the input c into the gene expression level g in the DTR as well as the ITR model can be either repressive or activating. We find no qualitative differences between the two regulation schemes, and thus present only the repressive scheme. The channel capacities of the two schemes, separately optimized, differ by ~ 0.1 bits, with activation performing slightly better at $C \lesssim 1$, and vice versa. While the shapes of the regulatory curves must be different by construction, the general trends—sharper regulatory curves already for low C in the ITR model, lower dissociation constants for high C in the ITR model—for the activators are identical to those of the repressors. Second, we examined how the precise description of the input noise form (Φ_B in Eq. (4)) influences our results. In one limit, we can set $\Phi_B = \Phi_1$, corresponding to a single binding site, while in the opposite limits we can set $\Phi_B = \Phi_\infty$, corresponding to a large number of binding sites [28, 34, 35]. Tracing through the full numerical analysis in both cases, all differences are small, and largely confined to the regime in which both C and M are small.

Our analytical model is made tractable by several technical assumptions: The input noise depends on diffusion constants, and we assumed equal diffusion constants for the input and the intermediary molecules, and further that translationally regulated mRNAs are immobile. Relaxing these assumptions changes the effective diffusion rates in the diffusion noise terms; in particular, highly mobile mRNAs might further contribute to input noise reduction in the ITR model, Eq. (20). We also had to make assumptions about how the translationally regulated mRNAs are produced and how they interact with the inputs. Apart from a change in ITR noise magnitude, the model where translationally regulated mRNAs are present at a fixed molecule count behaves identically to the model we studied, where mRNAs are continuously made and degraded. We have, however, not analyzed more complicated interaction schemes between the in-

put and the mRNA, e.g., one in which the mRNA would sequester the input regulatory proteins. Lastly, we have assumed that the noise in the system is Gaussian and not overly large. This is supported by experimental evidence for the paradigmatic case of *Drosophila* embryogenesis, where crucial protein signals appear to be tightly controlled both at the input and output side, with typical coefficients of variation $\lesssim 10\%$ [40, 77–79]. Extending the applicability of our model to systems in which the noise is non-Gaussian and/or large would require a separate major computational effort (to numerically compute the channel properties, find its capacity, and then optimize it over parameters), which is beyond the scope of this work.

We focused exclusively on intrinsic noise arising from intracellular gene production and diffusion and have neglected the effects of extrinsic noise. A recent study explored the combined impact of intrinsic and extrinsic noise within the framework of information theory, using stochastic simulations, and found that extrinsic noise can cause overestimation of the channel capacity and other nontrivial effects [80]. While studying extrinsic noise with similar detail in our model would require a completely new effort, we briefly assessed the influence of extrinsic noise in the mRNA number M in an approximate calculation (see Appendix B); our results indicate that for a typical magnitude of the variance in M, the capacity loss due to extrinsic noise in the input-noise dominated regime ($C < 1$) is small.

What are the limits to noise reduction using translational regulation? We started with a pedagogical example where the input concentration is detected by M regulatory sites. This scenario can be realized as a consequence of gene amplification, which is commonplace in bacteria, but the number of gene duplications (i.e., M) typically stays below 10 [81]. While gene amplification also occurs in some strongly specialized eukaryotic cells, it is generally seen as a strategy to massively increase the protein production rate [82–84], or as a hallmark of cancer [85, 86]. Alternatively, a high multiplicity of M regulatory sites could occur in a single promotor / enhancer region. Consequently, the sites would be packed very closely on the DNA and thus the molecule "detected" by one site would have a high likelihood of rebinding to a nearby site, providing a statistically correlated, rather than independent, sample of the concentration in the bulk. In this regime, the decrease in input noise variance would be slower than $1/M$, and as $M \to \infty$, the input noise would saturate to a bound determined by the linear dimension of the cluster of regulatory sites [2, 28, 70]. The same argument will ultimately apply to M mRNA molecules of the ITR scenario. There, each mRNA harbors a binding site of size $\ell_c \sim 1 - 10$ nm, but the mRNAs typically are separated by distances $d \sim 1 - 10$ μm. Roughly, we can expect that input readouts will be independent and the ITR mechanism effective so long as $M \lesssim d/\ell_c \sim 10^2 - 10^4$, putting M in the range that we examined and where it appears biologically plausible;

recent experiments in the *Drosophila* and zebrafish embryos and in mammalian cells reported mRNA counts in the range of $50 - 1200$ per typical cell volume [79, 87–91]. For such values of M, we find significant increases in capacity of $0.8 - 1.8$ bits at $C \sim 0.1 - 1$, with optimal regulatory curves using high Hill coefficients $|H^*| \gtrsim 4$, thereby accessing the full output dynamic range. Moreover, the molecular mechanism for integrating the readout over M such "receptors" is simple diffusion of the intermediary protein y, unlike in the pedagogical example of M regulatory sites on the DNA where we know of no plausible molecular integration mechanism.

Are there biological examples of such indirect translational regulation scheme? During early *Drosophila* morphogenesis, one of the primary transcription factors that drive the anterior-posterior (AP) patterning of the embryo is Bicoid, which is established in a graded, exponentially decaying profile along the AP axis [92–96]. Bicoid activates a number of downstream genes directly in the anterior and the middle region of the embryo where its concentration is highest. Absolute concentrations of bicoid are estimated to be ~ 55 nM at a maximum, falling to ~ 8 nM at the midpoint of the embryo [40, 77]). At the posterior end of the embryo, where its concentration is low and hard to detect quantitatively using the imaging methods available, Bicoid translationally represses *caudal* mRNA [46, 47, 97, 98]. The *caudal* mRNA molecules are produced by the mother, who deposits them into the egg with a uniform distribution along the AP axis. As they are bound and repressed by Bicoid, while the free ones get translated into Caudal protein, the embryo develops a new gradient of Caudal protein with high concentration at the posterior and low in the anterior. Caudal acts much like the intermediary protein y in our model, becoming a regulator of patterning genes in the posterior. While such a scheme involving an intermediary maternal mRNA and protein appears wasteful at first glance, here we have shown that it may provide a substantial benefit for information transmission at low input concentration, i.e. in posterior regions of the embryo. Although we are not familiar with any direct measurement of *caudal* mRNA copy numbers, the copy numbers for the gap genes [79] (and also *bicoid* [87]) are consistent with the range of $M \sim 10^2 - 10^3$ mRNA per nuclear volume.

More generally, it is interesting to note that Bicoid is just one member of a larger homeodomain protein family whose members have the ability to act both as a transcriptional as well as translational regulators [45–47]. Several other proteins acting as translational regulators have been reported in the fly embryo and oocyte [50–57], and in other eukaryotic species [58–60], oftentimes in a developmental context. These findings hint at the possibility that the regulatory principle implemented by Bicoid and *caudal* mRNA in *Drosophila* may be at work also in other developmental systems. While translational regulation has also been reported in bacteria [61–67], here typical mRNA numbers are of order ~ 1 [99], such that in prokaryotes translational regulation may serve a different purpose than averaging the input to enhance information transmission. In summary, our results suggest that the dual role of proteins as both transcriptional and translational regulators is not just an accident of history, but a solution to a fundamental physics problem faced by the cell.

Appendix A: Shot-noise propagation in a generic signaling cascade

Let us assume the following signaling cascade, described by a system of Langevin equations for the signaling species x_j, in which shot noise is generated at each level $j \in [1,..,n]$ and (for $j < n$) propagated into the copy number of species x_{j+1} via a regulatory function $f_{j+1}(x_j)$:

$$\frac{dx_1}{dt} = r_1 - \frac{1}{\tau_1}x_1 + \xi_1$$
$$\vdots$$
$$\frac{dx_n}{dt} = r_n f_n(x_{n-1}) - \frac{1}{\tau_n}x_n + \xi_n \ , \qquad (A1)$$

In the steady state, for the noise powers of the Langevin noise sources ξ_j we have

$$\langle \xi_j(t)\xi_k(t')\rangle = \left(R_j + \frac{\bar{x}_j}{\tau_j}\right)\delta_{jk}\delta(t-t') = \frac{2\bar{x}_j}{\tau_j}\delta_{jk}\delta(t-t') \ , \tag{A2}$$

where $R_1 \equiv r_1$ and $R_{j>1} \equiv r_j f_j(x_{j-1})$, and $\bar{x}_j = \langle x_j \rangle$ denotes the stationary mean. For our purposes we want to compute the overall variance in the last component $\sigma_n^2 = \langle \delta x_n^2 \rangle$.

Linearizing around the means $\bar{x}_j$ via $x_j = \bar{x}_j + \delta x_j$ and $f_j(x_{j-1}) \simeq f_j(\bar{x}_{j-1}) + f_j'(\bar{x}_{j-1})\delta x_{j-1}$, and Fourier-transforming for $t \to \omega$ allows us to convert Eqs. (A1) into the following equation system for the Fourier-transformed fluctuations $\delta \tilde{x}_j$:

$$-i\omega\delta\tilde{x}_1 = -\frac{1}{\tau_1}\delta\tilde{x}_1 + \tilde{\xi}_1$$
$$\vdots$$
$$-i\omega\delta\tilde{x}_n = r_n f_n'(\bar{x}_{n-1})\delta\tilde{x}_{n-1} - \frac{1}{\tau_n}\delta\tilde{x}_n + \tilde{\xi}_n \ . \quad (A3)$$

For the Fourier-transformed noise powers $\tilde{\xi}_j$ we get

$$\langle \tilde{\xi}_j(\omega)\tilde{\xi}_k^*(\omega)\rangle = 2\tau_j^{-1}\bar{x}_j\delta_{jk} \ . \qquad \text{(A4)}$$

The system defined by Eqs. (A3) can be solved algebraically by succesively inserting the solution for $\delta\tilde{x}_j$ into the equation for $\delta\tilde{x}_{j+1}$:

$$\delta\tilde{x}_1 = \frac{\tilde{\xi}_1}{\tau_1^{-1} - i\omega}$$

$$\vdots$$

$$\delta\tilde{x}_n = \sum_{j=1}^{n} \frac{\tilde{\xi}_j}{\tau_j^{-1} - i\omega} \prod_{q=j+1}^{n} \frac{r_q f_q'}{\tau_q^{-1} - i\omega} \ , \qquad \text{(A5)}$$

where we abbreviate $f_q' \equiv f_q'(\bar{x}_{q-1})$.

We can now obtain the variance σ_n^2 by integrating over the noise power spectrum $S_n(\omega) = \langle\delta\tilde{x}_n\delta\tilde{x}_n^*\rangle$ of the fluctuations in component n:

$$\sigma_n^2 = \int \frac{d\omega}{2\pi} S_n(\omega) = \int \frac{d\omega}{2\pi} \langle\delta\tilde{x}_n\delta\tilde{x}_n^*\rangle$$

$$= \int \frac{d\omega}{2\pi} \sum_{j=1}^{n} \frac{2\bar{x}_j\tau_j^{-1}}{\tau_j^{-2} + \omega^2} \prod_{q=j+1}^{n} \frac{\left(r_q f_q'\right)^2}{\tau_q^{-2} + \omega^2} \ , \qquad \text{(A6)}$$

where we recall that for mixed indices $\langle\tilde{\xi}_j(\omega)\tilde{\xi}_k^*(\omega)\rangle = 0$.

Isolating the n-term of the sum, and introducing the dimensionless integration variable $w \equiv \tau_n\omega$, we can further write:

$$\sigma_n^2 = \bar{x}_n \underbrace{\int \frac{dw}{\pi} \frac{1}{1+w^2}}_{=1}$$

$$+ \sum_{j=1}^{n-1} \int \frac{dw}{\pi} \frac{\bar{x}_j\left(\frac{\tau_j}{\tau_n}\right)}{1 + \left[\frac{\tau_j}{\tau_n}w\right]^2} \prod_{q=j+1}^{n} \frac{\left(r_q\tau_q f_q'\right)^2}{1 + \left[\frac{\tau_q}{\tau_n}w\right]^2} \ . \qquad \text{(A7)}$$

The integrand of the second integral can be written as:

$$\prod_{k=j}^{n} \left(1 + \left[\frac{\tau_k}{\tau_n}w\right]^2\right)^{-1} = \frac{1}{1+w^2} \prod_{k=j}^{n-1} \left(1 + \left[\frac{\tau_k}{\tau_n}w\right]^2\right)^{-1} \ . \qquad \text{(A8)}$$

The leading factor $(1+w^2)^{-1}$ only has significant contributions when $|w| \lesssim 1$, and will be suppressed (together with the whole integrand) for $|w| \gg 1$. Assuming that τ_n is the longest timescale of the problem, i.e. $\forall k : \tau_n \gg \tau_k$, in the relevant regime $|w| \lesssim 1$ all factors except for the leading one will be $\simeq 1$. Thus, to a good approximation:

$$\sigma_n^2 \simeq \bar{x}_n + \sum_{j=1}^{n-1} \bar{x}_j\left(\frac{\tau_j}{\tau_n}\right) \prod_{q=j+1}^{n} \left(r_q\tau_q f_q'\right)^2 \int \frac{dw}{\pi}\frac{1}{1+w^2}$$

$$= \bar{x}_n + \sum_{j=1}^{n-1} \underbrace{\left[\prod_{q=j+1}^{n} r_q\tau_q f_q'\right]^2 \times \left(\frac{\tau_j}{\tau_n}\right) \times \bar{x}_j}_{\equiv \sigma_{n\leftarrow j}^2} \ . \quad \text{(A9)}$$

Appendix B: Influence of extrinsic noise in the mRNA number M

Here we briefly explore the impact of extrinsic noise in the mRNA copy number M, the most important parameter of our ITR model. In the original model mRNA is constantly produced and degraded at finite rates, generating an intrinsic contribution to the overall noise. Here, in contrast, we assume that mRNAs are present at a fixed copy number M in each system, but that across systems, the value of M is normally distributed with a mean $\bar{M}$ and a variance σ_M^2. Note that this scenario corresponds to the limit of infinitely fast mRNA production and degradation in our original model, implying that the Fano factor $F \simeq 1$. If the typical deviations of M from the mean $\bar{M}$ are small, to a first approximation we can expand (dropping the subscript $P(M)$ for brevity)

$$\langle\sigma_c^2(M)\rangle = \langle\sigma_c^2(\bar{M})\rangle + \underbrace{\langle(\sigma_c^2)'_{\bar{M}}(M - \bar{M})\rangle}_{=0}$$

$$+ \langle(\sigma_c^2)''_{\bar{M}}(M - \bar{M})^2\rangle + \dots$$

$$\simeq \sigma_c^2(\bar{M}) + (\sigma_c^2)''_{\bar{M}}\ \sigma_M^2 \ , \qquad \text{(B1)}$$

where angular brackets denote averaging over the extrinsic noise distribution and the uneven terms vanish because this distribution is assumed to be Gaussian. The effective variance at the input, $\sigma_c^2(M)$, has the form [Eq. (29)]

$$\sigma_c^2(M) = \frac{1}{N_{\max}}\left(\mathcal{N} + \frac{1}{M}\mathcal{R}\right) \ , \qquad \text{(B2)}$$

where $\mathcal{N}$ corresponds to the output noise [term (i) in Eq. (29)] which is independent of M (note that $f' \sim 1/M$), and $\mathcal{R}$ collects all the remaining terms after isolating their common prefactor $1/M$ (note that the per-mRNA occupancy function $h(c)$ and its derivative do not depend on M). It follows that

$$(\sigma_c^2)''_{\bar{M}} = \frac{1}{N_{\max}}\frac{2}{\bar{M}^3}\mathcal{R} = \frac{2}{\bar{M}^2}\left(\sigma_c^2(\bar{M}) - \frac{1}{N_{\max}}\mathcal{N}\right) \ , \tag{B3}$$

and hence

$$\langle\sigma_c^2(M)\rangle \simeq \sigma_c^2(\bar{M}) + \frac{2\sigma_M^2}{\bar{M}^2}\left(\sigma_c^2(\bar{M}) - \frac{1}{N_{\max}}\mathcal{N}\right) \ . \quad \text{(B4)}$$

Let us now assume $\sigma_M^2 = \beta^2\bar{M}^\alpha$, which we can use to study two relevant example regimes: the case in which extrinsic noise has a fixed coefficient of variation (CV), $\sigma_M/\bar{M} = \beta$ ($\alpha = 2$), and the case in which the extrinsic noise is purely Poissonian, $\sigma_M^2 = \bar{M}$ ($\alpha = 1, \beta = 1$). We can then write:

$$\langle\sigma_c^2(M)\rangle \simeq \sigma_c^2(\bar{M}) + \frac{2\beta^2}{\bar{M}^{2-\alpha}}\left(\sigma_c^2(\bar{M}) - \frac{1}{N_{\max}}\mathcal{N}\right)$$

$$\leq \sigma_c^2(\bar{M})\left(1 + \frac{2\beta^2}{\bar{M}^{2-\alpha}}\right) \ . \qquad \text{(B5)}$$

As expected, fluctuations in M thus increase the effective variance in our model, but this increase is bounded by a factor $\psi = \left(1 + 2\beta^2/\bar{M}^{2-\alpha}\right)$; the corresponding (maximal) loss in capacity is given by an additive term $\Delta I = -\frac{1}{2}\log_2\psi$. For a typical fixed CV of 10%, $\psi \equiv (1 + 2\beta^2) = 1.02$. In the Poissonian noise case, $\psi \equiv (1 + 2/\bar{M})$, which quickly approaches 1 as $\bar{M}$ grows large; specifically, for $\bar{M} \geq 100$ we have $\psi \leq 1.02$. For $\psi = 1.02$ the expected loss in capacity is below 0.02 bits, small compared to the expected capacity increase in the input-noise dominated regime ($C < 1$) for $\bar{M} \gtrsim 10$.

[1] M. Ptashne, *A Genetic Switch: Gene Control and Phage Lambda* (Cell Press & Blackwell Scientific Publications, 1986), ISBN 0865423156.
[2] H. C. Berg and E. M. Purcell, Biophys J **20**, 193 (1977).
[3] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[4] G. Tkačik and A. M. Walczak, J Phys Condens Matter **23**, 153102 (2011).
[5] G. Tkačik and W. Bialek, arXiv:1412.8752 [q-bio.QM] (2014).
[6] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, 2012), ISBN 0691138915.
[7] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code (Computational Neuroscience)* (A Bradford Book, 1999), ISBN 0262681080.
[8] G. Tkačik, C. G. Callan, and W. Bialek, Proc Natl Acad Sci (USA) **105**, 12265 (2008).
[9] A. S. Hansen and E. K. O'Shea, eLife Sciences p. e06559 (2015).
[10] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, Proc Natl Acad Sci (USA) **110**, 16301 (2013).
[11] G. Tkačik, J. O. Dubuis, M. D. Petkova, and T. Gregor, Genetics **199**, 39 (2015).
[12] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, Science **334**, 354 (2011).
[13] J. Selimkhanov, B. Taylor, J. Yao, A. Pilko, J. Albeck, A. Hoffmann, L. Tsimring, and R. Wollman, Science **346**, 1370 (2014).
[14] E. Ziv, I. Nemenman, and C. H. Wiggins, PLoS ONE **2**, e1077 (2007).
[15] G. Tkačik, C. G. Callan, and W. Bialek, Phys Rev E **78**, 011910 (2008).
[16] G. Tkačik, A. M. Walczak, and W. Bialek, Phys Rev E **80**, 031920 (2009).
[17] A. M. Walczak, G. Tkačik, and W. Bialek, Phys Rev E **81**, 041905 (2010).
[18] G. Tkačik, A. M. Walczak, and W. Bialek, Phys Rev E **85**, 041903 (2012).
[19] T. R. Sokolowski and G. Tkačik, Phys Rev E **91**, 062710 (2015).
[20] G. Rieckh and G. Tkačik, Biophys J **106**, 1194 (2014).
[21] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys Rev E **82**, 031914 (2010).
[22] F. Tostevin and P. R. ten Wolde, Phys Rev Lett **102**, 218101 (2009).
[23] W. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys Rev Lett **107**, 048101 (2011).
[24] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys Rev E **86**, 021913 (2012).
[25] H. B. Barlow, in *Sensory Communication*, edited by W. A. Rosenbluth (MIT, Cambridge, MA, USA, 1961), pp. 217–234.
[26] G. Tkačik, T. Gregor, and W. Bialek, PLoS ONE **3**, e2774 (2008).
[27] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc Natl Acad Sci (USA) **99**, 12795 (2002).
[28] W. Bialek and S. Setayeshgar, Proc Natl Acad Sci (USA) **102**, 10040 (2005).
[29] W. Bialek and S. Setayeshgar, Phys Rev Lett **100**, 258101 (2008).
[30] G. Tkačik and W. Bialek, Phys Rev E **79**, 051901 (2009).
[31] R. G. Endres and N. S. Wingreen, Proc Natl Acad Sci (USA) **105**, 15749 (2008).
[32] R. G. Endres and N. S. Wingreen, Phys Rev Lett **103**, 158101 (2009).
[33] T. Mora and N. S. Wingreen, Phys Rev Lett **104**, 248101 (2010).
[34] K. Kaizu, W. de Ronde, J. Paijmans, K. Takahashi, F. Tostevin, and P. R. ten Wolde, Biophys J **106**, 976 (2014).
[35] J. Paijmans and P. R. ten Wolde, Phys Rev E **90**, 032708 (2014).
[36] P. R. ten Wolde, N. B. Becker, T. E. Ouldridge, and A. Mugler, arXiv:1505.06577 [physics, q-bio] (2015).
[37] B. A. Bicknell, P. Dayan, and G. J. Goodhill, Nat Commun **6**, 7468 (2015).
[38] R. Kafri, M. Levy, and Y. Pilpel, Proc Natl Acad Sci USA **103**, 11653 (2006).
[39] R. Kafri, M. Springer, and Y. Pilpel, Cell **136**, 389 (2009).
[40] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, Cell **130**, 153 (2007).
[41] T. Erdmann, M. Howard, and P. R. ten Wolde, Phys Rev Lett **103**, 258101 (2009).
[42] T. R. Sokolowski, T. Erdmann, and P. R. ten Wolde, PLoS Comput Biol **8**, e1002654 (2012).
[43] M. Tikhonov, S. C. Little, and T. Gregor, arXiv:1501.07342 [q-bio.MN] (2015).
[44] A. Mugler, A. Levchenko, and I. Nemenman, arXiv:1505.04346 [physics, q-bio] (2015).
[45] D. Niessing, W. Driever, F. Sprenger, H. Taubert, H. Jäckle, and R. Rivera-Pomar, Mol Cell **5**, 395 (2000).
[46] R. Rivera-Pomar, D. Niessing, U. Schmidt-Ott, W. J. Gehring, and H. Jäckle, Nature **379**, 746 (1996).
[47] J. Dubnau and G. Struhl, Nature **379**, 694 (1996).
[48] S. Iwasaki, T. Kawamata, and Y. Tomari, Mol Cell **34**, 58 (2009).
[49] S. Iwasaki and Y. Tomari, Fly (Austin) **3**, 204 (2009).
[50] O. Johnstone and P. Lasko, Annu Rev Genet **35**, 365 (2001).
[51] L. Chen, J. G. Dumelie, X. Li, M. H. Cheng, Z. Yang, J. D. Laver, N. U. Siddiqui, J. T. Westwood, Q. Morris, H. D. Lipshitz, et al., Genome Biol **15**, R4 (2014).
[52] M. R. Nelson, A. M. Leidal, and C. A. Smibert, EMBO

J **23**, 150 (2004).

[53] J. L. Semotok, R. L. Cooperstock, B. D. Pinder, H. K. Vari, H. D. Lipshitz, and C. A. Smibert, Curr Biol **15**, 284 (2005).

[54] B. D. Pinder and C. A. Smibert, EMBO Rep **14**, 80 (2012).

[55] B. D. Pinder and C. A. Smibert, Fly **7**, 142 (2013).

[56] A. Yarunin, R. E. Harris, M. P. Ashe, and H. L. Ashe, RNA Biol **8**, 904 (2011).

[57] M. Okabe, T. Imai, M. Kurusu, Y. Hiromi, and H. Okano, Nature **411**, 94 (2001).

[58] L. Paillard, F. Omilli, V. Legagneux, T. Bassez, D. Maniey, and H. Osborne, EMBO J **17**, 278 (1998).

[59] J. Delaunay, G. L. Mée, N. Ezzeddine, G. Labesse, C. Terzian, M. Capri, and O. Aït-Ahmed, Nucl Acids Res **32**, 3070 (2004).

[60] V. A. Marin and T. C. Evans, Development **130**, 2623 (2003).

[61] E. Van Assche, S. Van Puyvelde, J. Vanderleyden, and H. P. Steenackers, Front Microbiol **6**, 141 (2015).

[62] P. Babitzke, Curr Opin Microbiol **7**, 132 (2004).

[63] H. Yakhnin, H. Zhang, A. V. Yakhnin, and P. Babitzke, J Bacteriol **186**, 278 (2004).

[64] H. Yakhnin, A. V. Yakhnin, and P. Babitzke, Mol Microbiol **61**, 1252 (2006).

[65] E. Merino, P. Babitzke, and C. Yanofsky, J Bacteriol **177**, 6362 (1995).

[66] P. Gollnick, Mol Microbiol **11**, 991 (1994).

[67] Z. Liu, H. Wang, H. Wang, J. Wang, Y. Bi, X. Wang, R. Yang, and Y. Han, Front Microbiol **6**, 862 (2015).

[68] L. Giorgetti, T. Siggers, G. Tiana, G. Caprara, S. Notarbartolo, T. Corona, M. Pasparakis, P. Milani, M. L. Bulyk, and G. Natoli, Mol Cell **37**, 418 (2010).

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

[70] A. M. Berezhkovskii and A. Szabo, J Chem Phys **139** (2013).

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

[72] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Phys Rev E **72**, 051907 (2005).

[73] A. M. Walczak, M. Sasai, and P. G. Wolynes, Biophys J **88**, 828 (2005).

[74] A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Proc Natl Acad Sci (USA) **102**, 18926 (2005).

[75] A. Mugler, A. M. Walczak, and C. H. Wiggins, Phys Rev E **80**, 041921 (2009).

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

[77] T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank, Cell **130**, 141 (2007).

[78] J. O. Dubuis, R. Samanta, and T. Gregor, Mol Syst Biol **9**, 639 (2013).

[79] S. C. Little, M. Tikhonov, and T. Gregor, Cell **154**, 789 (2013).

[80] S. S. Mc Mahon, O. Lenive, S. Filippi, and M. P. H. Stumpf, J R Soc Interface **12**, 20150597 (2015).

[81] D. I. Andersson and D. Hughes, Annu Rev Genet **43**, 167 (2009).

[82] J. C. Kim, J. Nordman, F. Xie, H. Kashevsky, T. Eng, S. Li, D. M. MacAlpine, and T. L. Orr-Weaver, Genes Dev **25**, 1384 (2011).

[83] J. M. Claycomb, M. Benasutti, G. Bosco, D. D. Fenger, and T. L. Orr-Weaver, Dev Cell **6**, 145 (2004).

[84] A. C. Spradling and A. P. Mahowald, Proc Natl Acad Sci USA **77**, 1096 (1980).

[85] A. Matsui, T. Ihara, H. Suda, H. Mikami, and K. Semba, Biomol Concepts **4**, 567 (2013).

[86] D. G. Albertson, Trends Genet **22**, 447 (2006).

[87] S. C. Little, G. Tkačik, T. B. Kneeland, E. F. Wieschaus, and T. Gregor, PLoS Biol **9**, e1000596 (2011).

[88] M. D. Petkova, S. C. Little, F. Liu, and T. Gregor, Current Biology **24**, 1283 (2014).

[89] Y. Oka and T. N. Sato, Sci Rep **5**, 8571 (2015).

[90] K. Bahar Halpern, S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg, and S. Itzkovitz, Mol Cell **58**, 147 (2015).

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

[92] W. Driever and C. Nüsslein-Volhard, Cell **54**, 83 (1988).

[93] W. Driever and C. Nüsslein-Volhard, Cell **54**, 95 (1988).

[94] P. A. Lawrence, *The Making of a Fly: The Genetics of Animal Design* (Wiley-Blackwell, 1992), ISBN 0632030488.

[95] S. F. Gilbert, *Developmental Biology* (Sinauer Associates Inc.,U.S., 2013), ISBN 1605351733.

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

[97] M. Mlodzik and W. J. Gehring, Cell **48**, 465 (1987).

[98] P. M. Macdonald and G. Struhl, Nature **324**, 537 (1986).

[99] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science **329**, 533 (2010).