entropy

MDPI

Article

Dissipation in Non-Steady State Regulatory Circuits

Paulina Szymańska-Rożek [1], Dario Villamaina [2,3], Jacek Miękisz [1] and Aleksandra M. Walczak [2,*]

[1] Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, 02-097 Warszawa, Poland; p.szymanska@gmail.com (P.S.-R.); miekisz@mimuw.edu.pl (J.M.)
[2] Laboratoire de Physique de l'École Normale Supérieure (PSL University), CNRS, Sorbonne Université, and Université de Paris, 75005 Paris, France; villamaina@gmail.com
[3] Capital Fund Management, 23 rue de l'Université, 75007 Paris, France
* Correspondence: awalczak@lpt.ens.fr

Received: 8 November 2019; Accepted: 5 December 2019; Published: 10 December 2019

Abstract: In order to respond to environmental signals, cells often use small molecular circuits to transmit information about their surroundings. Recently, motivated by specific examples in signaling and gene regulation, a body of work has focused on the properties of circuits that function out of equilibrium and dissipate energy. We briefly review the probabilistic measures of information and dissipation and use simple models to discuss and illustrate trade-offs between information and dissipation in biological circuits. We find that circuits with non-steady state initial conditions can transmit more information at small readout delays than steady state circuits. The dissipative cost of this additional information proves marginal compared to the steady state dissipation. Feedback does not significantly increase the transmitted information for out of steady state circuits but does decrease dissipative costs. Lastly, we discuss the case of bursty gene regulatory circuits that, even in the fast switching limit, function out of equilibrium.

Keywords: biomolecular circuits; regulation; dissipation; stochastic systems

1. Introduction

Cells rely on molecular signals to inform themselves about their surroundings and their own internal state [1]. These signals can describe the surrounding sugar type and concentration, which is the case of many bacterial operons, such as those used for lactose or galactose breakdown [2]. Signaling and activation of phosphorylated receptors provide a means of informing bacterial cells on faster timescales about a wide range of conditions including crowding, growth signals, and stress [3]. Triggered by these signals cells activate regulatory networks and cascades that allow them to respond in an appropriate way to existing signals.

A response is usually trigerred by a change in the environment, which perturbs the previous state of the cell and the regulatory system. Specifically, if the regulatory circuit was functioning in steady state, a change in the concentration of the signaling molecule, or the appearance of a new molecule will kick it out of steady state. Here we investigate the response to such perturbations.

In this paper we study abstract mathematical models whose goal is to capture the main regulatory features of biochemical circuits. Our models do not capture many of the details of biochemical complexity of regulatory units in real cells. By "circuit" or "network" throughout the paper we mean a set of stochastic processes that transform an input signal through a regulatory function to produce an output response. This use of the word "circuit" or network is standard in the biophysics literature [1–3]. Abstract models of biochemical circuits have proven useful in understanding molecular regulation in many biological systems from development to immunology [2–28].

The energy dissipated in a regulatory network comes on one hand from the fact that certain steps, for example producing proteins, require ATP. However, energy dissipation also measures how far out of equilibrium a given circuit functions by identifying irreversible (so ATP consuming) reactions [29–31].

Regulatory circuits that function out of equilibrium do not obey detailed balanced, which means they dissipate energy, even if they produce the same amount of proteins as circuits that function in equilibrium. We are interested in exploring the constraints that energy dissipation imposes on the types of regulatory functions. The motivation is not because of limiting energetic resources in cells; ATP is typically assumed to be abundant [32,33] or can be generated by burning carbon present in the cell. Rather we consider energy dissipation as a measure of irreversibility that allows us to compare the irreversibility of signaling encoded in given regulatory functions.

In order to concentrate on this specific problem of dissipation coming from regulatory functions, we choose to study a simplified model with two binary elements: a receptor and a protein. Each element can be in one of two states: active or inactive, and its state regulates the state of the other element. The first element—the receptor—is our input that responds to changes in the environment, and the second element—the regulatory protein such as a kinase in a two component signaling cascade—is the output of our regulatory system. We do not take into account the ATP-ADP balance for these reactions, but concentrate on the dissipation coming from the regulatory computation. Effectively, we assume that while ATP is certainly needed, it is part of the hardware of the network. In turn, we are interested in the question of given a certain set of hardware, what are the best regulatory functions (software) we can implement. We have reduced a description of a biochemical circuit to stochastic processes governing the flipping of binary variables and we will study the parameters of these processes.

Dissipation in molecular regulatory networks has received a lot of theoretical attention [29–31,34–38]. This line of thought goes back to the non-equilibrium scheme of kinetic proofreading [4,5] in which energy is used for error correction of the signal. A more recent application [29] has shown that energy dissipation is also needed for regulatory circuits to adapt to external signals and respond accurately. A similar conclusion that energy dissipation is necessary was reached for molecular circuits that try to learn about external concentrations [30] and it was shown that the amount of dissipated energy limits reliable readout [30,36,39–42]. Results linking information, dissipation and learning [43–45] have been derived in the general framework of stochastic thermodynamics [34,46]. In the context of biochemical reactions, both continuous biochemical kinetics models [30,41,42,47] and bipartite two state systems [40,43,48–51] have been used in this context. Among other topics the link between dissipation and prediction has been explored, again showing that long term prediction requires energy expenditure [35,47], and the non-predictive part of the information about past fluctuations is linked to dissipation [35]. Most recently, the links between information and dissipation have been studied in spatial systems [52].

A regulatory circuit fulfils a function and we assume that the goal of our network is to maximally transmit information between the input and output [12]. This objective function has been studied before theoretically, using both binary and more detailed models [53–59]. Others have also optimized the rate of information transmission [19–21,60]. Information transmission in regulatory circuits has also been investigated experimentally in fly development [18,61,62], NFκB signaling [28], calcium signaling [63] and dynamical readouts were compared to static information transmission between the input and output of ERK, calcium and NFκB signaling networks [64]. While it is an arbitrary choice of the objective function for a regulatory network, and many networks do not optimize information transmission, it is rather unlikely that a circuit aimed at sensing and responding to the environment does not transmit any information about the signal to the output. This choice of the objective function allows us to perform concrete calculations and investigate the trade-off between information and dissipation, which are both tied to the logic of the regulatory system.

Here, inspired by receptor-ligand binding, we use a simple two state system to build intuition about the trade-offs in information transmission, dissipation and functioning out of steady state. In a pedagogical spirit we remind the reader of the notions of information (Section 3), dissipation (Section 4)

and review some of our previous results from work that studied information transmission [65] and the trade-offs between information transmission and dissipation for regulatory circuits in steady state [66] (Sections 6.1 and 6.2). A signal often perturbs the system out of steady state, to which it then relaxes back. In this paper we calculate the non-equilibrium dissipation for circuits that function out of steady state and maximally transmit information between the input and a potentially delayed output given constraints on dissipation. While the setup of the optimization problem (Equation (13)) is the same as in our previous work [66], considering average dissipation (Equation (14)) is new (Sections 6.3 and 6.4).

Lastly, we include some comments on dissipation in simple gene regulatory circuits with bursty transcription (Section 7) [67–73]. We show how even a fast switching gene promoter need not be in equilibrium. Our goal is not to provide an exhaustive review of the field but to illustrate with simple examples some trade-offs that appear in these molecular circuits.

2. Model

We consider a system at time t consisting of two discreet random variables z_t and x_t, which describe the input state and output state of the systems, respectively. We previously used these abstract stochastic processes to study regulation in biochemical circuits in Mancini et al. [66] and such binary models of biochemical circuits have been studied by others [40,43,48–51]. For simplicity we assume that x and z can take only two values: $+$ (active state) and $-$ (inactive state). The input state corresponds to the presence or absence of a signaling molecule (or a high or low concentrations of a signaling molecule), whereas the output state is activation or not of a response pathway or regulator. The specific regulatory interactions between them will be defined later within the specfic studied model(s). At every time t, the system is in one of four possible states (z_t, x_t): $(-,-)$, $(-,+)$, $(+,-)$, or $(+,+)$. The master equation for the temporal evolution of the conditional probability distribution $p(z_t, x_t|z_0, x_0)$ of the system is:

$$\frac{\partial}{\partial t} p(z_t, x_t|z_0, x_0) = -\mathcal{L} p(z_t, x_t|z_0, x_0), \tag{1}$$

where $\mathcal{L}$ is a 4×4 matrix with transition rates between the four states. We will be interested in the joint probability $p(x_t, z_0)$, that is we will look at the output variable x at time t and the initial state of the input variable z:

$$p(x_t, z_0) = \sum_{x_0, z_t = \pm 1} p(z_t, x_t|z_0, x_0) \cdot p(x_0, z_0). \tag{2}$$

This probability is needed in the computation of the central quantity we optimize: the time-delayed mutual information between the initial state of the input and the state of the output at t (defined in Section 3). After marginalization over possible states of z_0 we will obtain $p(x_t) = \sum_{z_0} p(x_t, z_0)$, which in turn is indispensable for calculating the dissipation of the system defined in Section 4.

We restrict our analysis to symmetric models, in which we do not distinguish between the $(-,-)$ and $(+,+)$ states, and, analogously, between the $(-,+)$ and $(+,-)$ states. This is a simplification that is not motivated by a biological observation. The symmetry of the model allows us to write the probability distribution at any time t as $p(x_t, z_0) = \left(\frac{1+\mu_t}{4}, \frac{1-\mu_t}{4}, \frac{1-\mu_t}{4}, \frac{1+\mu_t}{4}\right)$, assuming the initial probability distribution also assumes the same symmetry: $p(x_0, z_0) = p_0 = \left(\frac{1+\mu_0}{4}, \frac{1-\mu_0}{4}, \frac{1-\mu_0}{4}, \frac{1+\mu_0}{4}\right)$. For the models in which the initial distribution is the steady state one, $p^{\text{init}} = p(x_0, z_0) = p^{\text{ss}}$, which imposes a condition on μ_0.

2.1. Model without Feedback: S and $\tilde{S}$

The first, simplest model we analyze is a symmetric model in which only the input affects the output and there is no feedback from the output to the input. The output variable either aligns

or anti-aligns to the input variable with rate r, regardless of the state of the input (see Figure 1A). The input variable z flips between active and inactive states with rate u and the output variable x aligns with rate r and anti-aligns with rate s (see Figure 1). The dynamics is given by a transition rate matrix given in Appendix A.

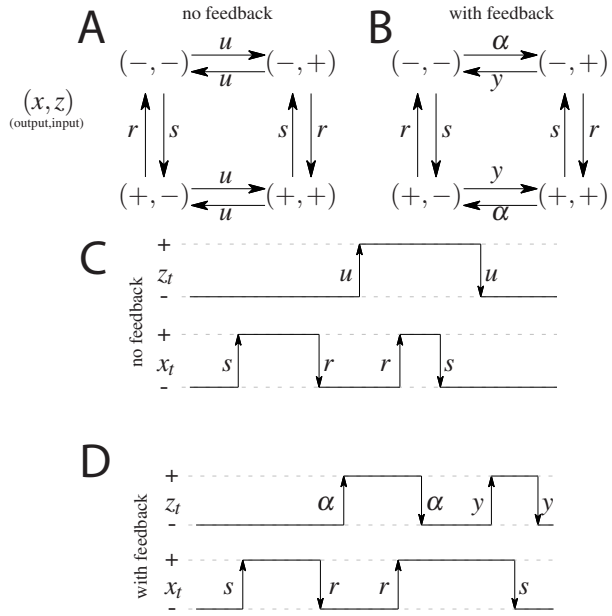

Figure 1. A cartoon of the possible states and transitions for both models: without feedback (**A**), and with feedback (**B**). Since there are two binary variables there are four states; transition rates are marked next to respective arrows. Note the symmetry between the "pure" $((-,-)$ and $(+,+))$ states and the "mixed" states $((-,+)$ and $(+,-))$ in both models. Representation of a possible time evolution of the system. Two variables flip between active $(+)$ and inactive $(-)$ states with respective rates. In the model without feedback (**C**) the output variable depends on the input variable (the output aligns to the input with rate r or anti-aligns, with rate s), the input variable z flips freely between its active and inactive state, regardless of the state of the output. In the model with feedback (**D**), there is a difference in rates of flipping of the input that depends on the state of the output.

We calculate analytically the joint probability distribution $p(x_t, z_0)$ (a four-dimensional vector) and marginal probability distributions $p(x_t)$ and $p(z_0)$ (two-dimensional random vectors), needed to find the mutual information, that we will define in Equation (3), as a function of the transition rates u, s, r, and a parameter μ_0 that parametrizes the initial state of the system (see Appendix B). We set, without loss of generality, one rate equal to 1, specifically $r = 1$. The specific expressions for the probability distributions for the occupancy of the four states for the model without feedback are given in Appendix A. In steady state the probability distribution for the occupancy of the four states simplifies to $p^\infty = \left(\frac{u+1}{2s+4u+2}, \frac{s+u}{2s+4u+2}, \frac{s+u}{2s+4u+2}, \frac{u+1}{2s+4u+2}\right)$.

We will consider this model in steady state, and we will call it model S. We will also allow for the initial conditions to be out of steady state, and then we will call it model $\tilde{S}$.

2.2. Models with Feedback: F and $\tilde{F}$

In the second analyzed model we allow the input variable to be dependent on the output, i.e., we allow for a feedback from x to z. We keep as much symmetry as possible, while still not distinguishing between the states $(-,-)$ and $(+,+)$, and between $(-,+)$ and $(+,-)$. The scheme is given in Figure 1B. In terms of the rates we allow the original input z_t switching parameters, to be different depending on the state of the output x_t introducing the rate α for anti-aligning the two variables and y for aligning the two variables. The notion of input and output is no longer meaningful since both variables influence each other. We note that this scheme is not the most general model

possible since we impose the symmetry between the 'pure' states, i.e., $(-,-)$ and $(+,+)$, and the 'mixed' states, i.e., $(-,+)$ and $(+,-)$, which reduces the number of parameters from 8 (as was studied in Mancini et al. [65]) to 4 (as was considered in Mancini et al. [66]). The transition matrix for this model, and the steady state probabilities are given in Appendix B.

Similarly to the case of the model without feedback, we consider the model with feedback in steady state and call it model F, or let the initial conditions be out of steady state by considering all values of μ_0 (model $\tilde{S}$). To sumarize, we use the following notation:

- S - no feedback, stationary initial condition;
- $\tilde{S}$ - no feedback, optimal initial condition;
- F - with feedback, stationary initial condition;
- $\tilde{F}$ - with feedback, optimal initial condition.

3. Information

The mutual information measured between the input z at time 0 and output x at time t is defined as [53,74]:

$$I[x_t, z_0] = \sum_{x_t, z_0} p(x_t, z_0) \log \frac{p(x_t, z_0)}{p(x_t)p(z_0)}. \tag{3}$$

In order to analyse the system in its natural timescale, we set $t = \tau/\lambda$, where λ is the inverse of the relaxation time (smallest, non-zero eigenvalue of the matrix $\mathcal{L}$) and calculate $I[x_\tau; z_0] = I[x_{\lambda \cdot t}; z_0]$. The term under the logarithm, which has been called the thermodynamic coupling function for systems with many degrees of freedom [75,76], describes the degree of correlation of the two variables, and is zero if the joined probability distribution factorizes. The thermodynamic coupling function has been shown to be useful to quantify the contributions of specific energy terms in binary models of allosteric systems [75,76].

Again exploiting the symmetry of the problem, the mutual information can be written as

$$I[x_t, z_0] = \frac{1}{2}\left((1+\mu)\log(1+\mu) + (1-\mu)\log(1-\mu)\right), \tag{4}$$

where $|\mu| \leq 1$. Since we have fixed $r = 1$, the symmetry of clockwise and counter-clockwise rotations is broken and $\mu \in [0, 1]$. Information is an increasing function of μ and is maximized at $I[x_t, z_0] = 1$ bit for $\mu = 1$. The specific values for μ are given in Appendixes A and B for the models with and without feedback.

4. Non-Equilibrium Dissipation

We consider the limitations on the regulatory functions coming from having a fixed amount of energy to dissipate during the signaling process that transmits information. Large amounts of dissipated energy allow systems to function far out of equilibrium, whereas no dissipated energy corresponds to equlibrium circuits. We quantify the degree to which the system functions out of equilibrium by comparing the probability of a forward, $P_{\rightarrow}(\vec{x})$, and backward, $P_{\leftarrow}(\vec{\tilde{x}})$, trajectory along the same path [34,77]:

$$\sigma = \sum_{\vec{x}} P_{\rightarrow}(\vec{x}) \log \frac{P_{\rightarrow}(\vec{x})}{P_{\leftarrow}(\vec{\tilde{x}})}, \tag{5}$$

where the paths are defined as $\vec{x} = (x_1, x_2, \ldots, x_N)$ and $\vec{\tilde{x}} = (x_N, x_{N-1}, \ldots, x_1)$ and each state x_i is a four dimensional probability of the input and output at time i. Using the Markov nature of the transitions $P(x_{t+1}|x_t)$ we write the probability of the forward path starting from the initial state x_1 as

$$P_{\rightarrow}(\vec{x}) = P_1(x_1) \prod_{t=1}^{N-1} P_{t \rightarrow t+1}(x_{t+1}|x_t), \tag{6}$$

and analogously for the backward path. Equation (5) now becomes:

$$\begin{aligned}\sigma = & \sum_{\vec{x}} P_{\rightarrow}(x_1, ..., x_{\mathrm{N}}) \log \frac{P_1(x_1) \prod_{t=1}^{\mathrm{N}-1} P_{t\rightarrow t+1}(x_{t+1}|x_t)}{P_N(x_N) \prod_{t=1}^{\mathrm{N}-1} P_{t+1\rightarrow t}(x_t|x_{t+1})} \\ = & \sum_{\vec{x}} P_{\rightarrow}(x_1, ..., x_{\mathrm{N}}) \log \frac{\prod_{t=1}^{\mathrm{N}-1} P_{t\rightarrow t+1}(x_{t+1}|x_t) P_t(x_t)}{\prod_{t=1}^{\mathrm{N}-1} P_{t+1\rightarrow t}(x_t|x_{t+1}) P_t(x_{t+1})},\end{aligned} \tag{7}$$

where we multiplied both the numerator and the denominator by the same product of probabilities $P(x_2) \cdot ... \cdot P(x_N)$. Simplifying further and marginalizing over the elements of $\vec{x}$ not equal to x_t or x_{t+1}:

$$\begin{aligned}\sigma &= \sum_{t=1}^{\mathrm{N}-1} \sum_{\vec{x}} P_{\rightarrow}(x_1, ..., x_{\mathrm{N}}) \log \frac{P_{t\rightarrow t+1}(x_{t+1}|x_t) P_t(x_t)}{P_{t+1\rightarrow t}(x_t|x_{t+1}) P_t(x_{t+1})} \\ &= \sum_{t=1}^{\mathrm{N}-1} P_{\rightarrow}(x_t, x_{t+1}) \log \frac{P_{t\rightarrow t+1}(x_{t+1}|x_t) P_t(x_t)}{P_{t+1\rightarrow t}(x_t|x_{t+1}) P_t(x_{t+1})} \\ &= \sum_t \sigma(t),\end{aligned} \tag{8}$$

which defines the time dependent dissipation production rate, $\sigma(t)$.

Noting that $P_{t\rightarrow t+1}(x_{t+1}|x_t) = P_{t+1\rightarrow t}(x_t|x_{t+1}) = P(x_{t+1} = i|x_t = j)$ and by explicitly defining the transition rates:

$$P(x_{t+1} = i|x_t = j) = w_{ij}\delta t + (1 - w_{ij}\delta t)\delta_{ij}, \tag{9}$$

and renaming $P_t(x_{t+1}) = p_j(t)$ and $P_y(x_t) = p_i(t)$ we obtain [34,77,78]:

$$\sigma(t) = \sum_{i,j} w_{ij} p_j(t) \log \frac{w_{ij} p_j(t)}{w_{ji} p_i(t)}, \tag{10}$$

that in the limit of $t \rightarrow \infty$ results in the *steady state entropy dissipation rate*:

$$\sigma^{\mathrm{ss}} = \sum_{i,j} p_j^{\mathrm{ss}} w_{ij} \log \frac{w_{ij}}{w_{ji}}, \tag{11}$$

where p_j^{ss} is the steady state probability distribution. We describe an alternative derivation of dissipation in Appendix C.

Again, we rescale the time in the above quantities by setting $t = \tau/\lambda$ (λ being is the inverse of the relaxation time):

$$\hat{\sigma}(\tau) = \frac{1}{\lambda}\sigma(\tau/\lambda), \qquad \hat{\sigma}^{\mathrm{ss}} = \frac{1}{\lambda}\sigma^{\mathrm{ss}}. \tag{12}$$

5. Setup of the Optimization

With these definitions, following Mancini et al. [66], we can ask what are the circuits that optimally transmit information given a limited constrained amount of steady state dissipation $\hat{\sigma}^{\mathrm{ss}}$:

$$\max_{\mathcal{L}|\hat{\sigma}^{\mathrm{ss}}} \mathcal{I}(\tau), \tag{13}$$

over the circuit's reaction rates, $\mathcal{L}$. The energy expense of a circuit that remains in steady state is well defined by this quantity. However the total expense of circuits that function out of steady state must be calculated as the integral of the entropy dissipation rate in Equation (10) over the entire time the circuit is active, τ_p, such as the duration of the cell cycle or the interval between new inputs that kick the system into the initial non-equilibrium state. After some time the circuit will relax to steady state (see the diagram in Figure 2) and its energetic expense is well described by the steady state dissipation. But the initial non-equilibrium steady state costs the system some energy. We can

compare the performance of circuits with different regulatory designs by considering the average energy expenditure until a given time τ_p:

$$\Sigma^{\text{avg}}(\tau_p) = \frac{1}{\tau_p}\int_0^{\tau_p} \hat{\sigma}(\tau)d\tau. \tag{14}$$

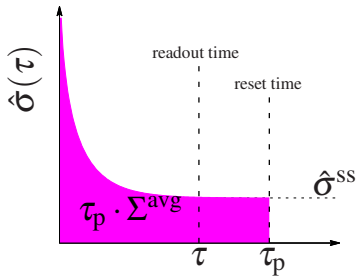

Figure 2. Schematic representation of system's relaxation. The entropy dissipation rate, $\hat{\sigma}(\tau)$ relaxes with time to its steady state value, $\hat{\sigma}^{ss}$. At τ_p the system is "kicked out" or reset, thus the pink area represents the total energy dissipated until that time. The information is collected at an earlier readout time τ.

We can foresee that circuits that spend most of their time in steady state will have their expenditure dominated by σ^{ss}, whereas circuits that spend a lot of time relaxing to steady state will be dominated by the additional out of steady state dissipation cost $\Delta\Sigma = \Sigma^{\text{avg}} - \hat{\sigma}^{ss}$. When $\tau_p \to \infty$, all circuits spend most of their time in steady state and the average integral in (14) converges to $\hat{\sigma}(\tau) \to \hat{\sigma}^{ss}$ as $\tau \to \infty$, so that the cost is dominated by the steady state dissipation.

Using the steady state distribution for model S and Equation (11) we can evaluate the non-rescaled steady state dissipation calculated for the model without feedback [66]

$$\sigma^{ss}(u,s) = \frac{(s-1)u\log_2(s)}{1+s+2u}. \tag{15}$$

If we impose a non-equlibrium state by setting $s \to 0$, the dissipation rescaled by the characteristic decay time (the lowest non-zero eigenvalue given by the minimum of the two non-zero eigenvalues $1+s$, and $2u$) tends to infinity

$$\hat{\sigma}^{ss}(u,s) = \sigma^{ss}/\lambda = \frac{(s-1)u\log_2(s)}{(1+s+2u)\cdot\min(1+s,2u)} \xrightarrow[s\to 0]{} \infty, \tag{16}$$

as expected. We also verify numerically that even in a non-steady state system that is kept out of equilibrium (Equation (10)) the rescaled dissipation (Equation (16)) tends to infinity, $\hat{\sigma} = \infty$ as $s \to 0$, for all τ, μ_0 and u.

The steady state dissipation rescaled by the smallest eigenvalue for models F and $\tilde{F}$ is [66]:

$$\hat{\sigma}^{ss}(\alpha,s,y) = \frac{2(\alpha - sy)}{A(A-\rho)}\log_2\left(\frac{\alpha}{sy}\right), \tag{17}$$

where

$$A = 1+s+y+\alpha, \tag{18}$$

$$\rho = \sqrt{(1+s+y+\alpha)^2 - 8(sy+\alpha)}. \tag{19}$$

6. Results

The task is to find maximal mutual information between the input and the output, with or without constraints, for all model variants, (regulation with and without feedback; starting at steady state, or starting out of steady state) and compare their performance—the amount of information transmitted and the energy dissipated. To build intuition we first summarize the results of the unconstrained optimization obtained by Mancini et al. [65]. Then, a constraint will be set on the steady state dissipation rate $\hat{\sigma}^{\rm ss}$ as in Mancini et al. [66]. We extend the latter results to models $\tilde{S}$ and $\tilde{F}$ by performing the optimization also with respect to the initial distribution. Finally, to compare not only the information transmitted in the models, but also its cost, we will calculate the average dissipation of the models.

In all cases we are looking for the maximum mutual information between the input at time 0 and the output at time τ, in the space of parameters (u, s and r for the model without feedback and α, y, s and r for the model with feedback). We can also treat the initial distribution (parametrized by a single parameter, μ_0), as an additional constraint or set μ_0 to be equal to $\mu_0^{\rm ss}$, i.e., fix the initial distribution to be the steady state one. Optimizing with a constraint is looking for the maximum of the function not in the whole parameter space ($\mathbb{R}_+^{\mathbb{N}}$), but on the manifold given by $\sigma^{\rm ss}$(parameters) = constraint. Finally, to compare not only the information transmitted in the models, but also its cost, we will calculate the average dissipation of the models.

6.1. Unconstrained Optimization

The results of the unconstrained optimization are summarized in Figure 3. As expected the maximum amount of information that can be transmitted decays with the readout time for all models. Feedback allows for better information transmission only in the case when the initial distribution is fixed to its steady state value. Optimizing over the initial distribution renders the models considered here without ($\tilde{F}$) and with feedback ($\tilde{S}$) equivalent. In this case the system relies on its initial condition and information loss is due to the system decorrelating and loosing information about its initial state. For a fixed initial distribution the model with feedback performs better than the model without feedback. We note that the feedback model considered here is a simplified model compared to the one studied in Mancini et al. [65], with less parameters. A full asymmetric model with feedback can transmit more information than a model without feedback if the initial conditions are not in steady state. However these architerctures correspond to infinite dissipation solutions since all backward rates are forbidden and the circuit can never regain its initial state since one of the states i becomes absorbing, $p_\infty(y') = \delta_{y',i}$, and attracts the whole probability weight. We are therefore restricting our exploration of models with feedback to the subclass without an absorbing steady state.

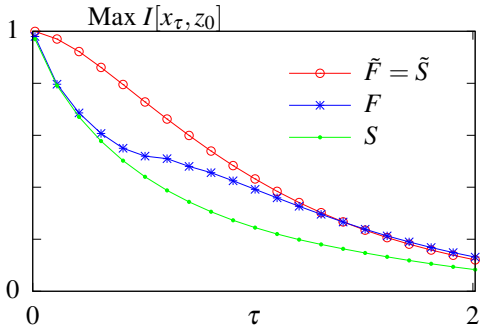

Figure 3. Results of the unconstrained optimization—mutual information for the models without feedback (S and $\tilde{S}$) and with feedback (F and $\tilde{F}$) with respect to the readout time τ. Optimization done both when the initial distribution is fixed to its steady state value (no tilde) and when the parameter is subjected to optimization as well (with tilde).

The modes of regulation of the circuits corresponding to the optimal solutions were discussed in previous work [65,66]. In short, the information-optimal steady state system uses rates that break the detailed balance and induce an order in visiting the four states i. Feedback increases the transmitted information for long time delays by implementing these cycling solutions using a mixture of fast and slow rates. Allowing for out of steady state initial conditions, circuits relax to absorbing final states that need to be externally reset. In this case the optimal solution with and without feedback both result in the stochastic processes cycling through the four states and simply relies on the decorrelation of the initial state.

6.2. Constraining $\hat{\sigma}^{ss}$

We next looked for rates that maximize the transmitted information $I[x_\tau, z_0]$ at a fixed time τ given a fixed steady state dissipation rate $\hat{\sigma}^{\rm ss}$. We first plot the maximal mutual information as function of the readout time, τ, for models without feedback, S (dashed lines) and $\tilde{S}$ (solid lines), (Figure 4). Not surprisingly, maximum information is a decreasing function of τ for both models, larger values of steady state dissipation, $\hat{\sigma}^{\rm ss}$, allow for more information transmitted, and model $\tilde{S}$ with optimized initial conditions transmits more information than model S, which remains in steady state.

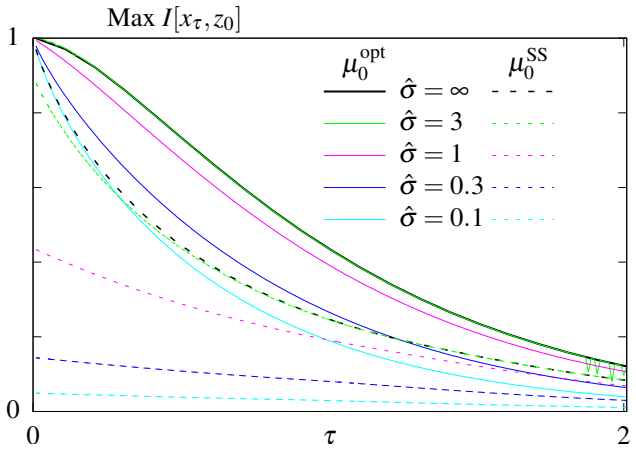

Figure 4. Results of the optimization problem with constrained steady state dissipation for models without feedback. Optimal mutual information as function of the readout time, τ, for different constrained steady state dissipation rates, $\hat{\sigma}^{\rm ss}$, for the model S (dashed lines) an $\tilde{S}$ (solid lines).

However, comparing all four models, the conclusion about the equivalence of the out of steady state model with ($\tilde{F}$) and without ($\tilde{S}$) feedback no longer holds when we constrain $\hat{\sigma}^{\rm ss}$ (Figure 5). The difference between optimal mutual information transmitted in models $\tilde{S}$ and $\tilde{F}$ is higher for systems that have smaller dissipation budgets $\hat{\sigma}^{\rm ss}$, and, as shown previously (Figure 3), the difference vanishes as $\hat{\sigma}^{\rm ss} \to \infty$. The remaining conclusions from Figure 4 hold: models with feedback transmits more information than models without feedback and models with free initial distributions transmit more information than the steady state models, as in the unconstrained optimization case (Figure 3).

Phase diagrams describing the optimal modes of regulation for steady state circuits are reported in Mancini et al. [66]. At large dissipation rates, the optimal out-of-equilibrium cicruits exploit the increased decorrelation time of the system since cycling solutions are permitted. Close to equilibrium, circuits with no feedback cannot transmit a lot of information. Circuits with feedback use a combination of slow and fast rates to transmit information. The optimal close to equilibrium regulatory functions rapidly align the two variables z_t and x_t ($y > \alpha$, s small), and slowly anti-aligns them, increasing the probability to be in the aligned $(+,+)$ and $(-,-)$ states. This results in a positive feedback loop. The same strategy of adjusting rates is used far from equilibrium but this time results in a cycling solution which translated into a negative feedback loop ($\alpha > y$, $s \approx 0$).

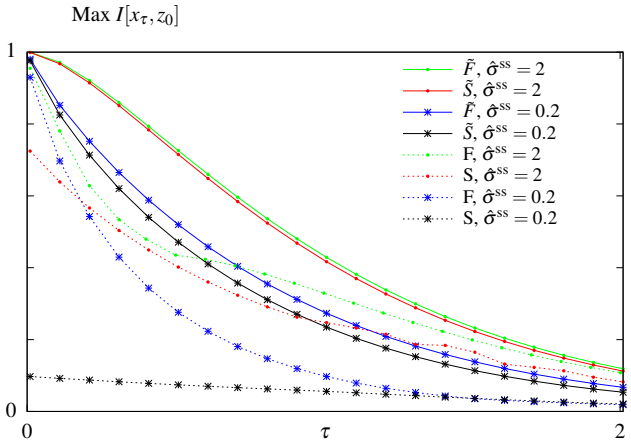

Figure 5. Results of the optimization problem with constrained steady state dissipation for all four models. Optimal mutual information as function of the readout time, τ, for two different constrained steady state dissipation rates, $\hat{\sigma}^{ss}$, for the models S and F (dashed lines), and the models $\tilde{S}$ and $\tilde{F}$ (solid lines).

Allowing the circuit to function out of steady state optimizes the initial condition μ_0 to be as far as possible from the steady state. The optimal initial condition is $\mu_0 = 1$, where only the aligned states are occupied (the initial distribution is $p_0 = (0.5, 0, 0, 0.5)$). This initial condition combined with $u < r$ and $s < r$ (Figure A1) decreases the decorrelation time and even a circuit with no feedback can transmit non-zero information. The rates of the circuits without feedback are simply set by the dissipation constraint, with $s \to 0$ for large dissipation and taking the value to balance u close to equilibrium (Figure A1). Optimal circuits far from equilibrium were reported in Mancini et al. [66] and close to equilibrium are shown in Figure 6. Circuits with feedback also mostly rely on the decorrelation of the initial state. Since the majority of the initial probability weight is in the aligned states, the y and α are always roughly equal (Figure A2). Only at intermediate dissipation rates, y slightly smaller than α and small s stabliize the initial aligned states and further decrease the decorrelation time (Figure 6), encoding small negative feedback in the circuit.

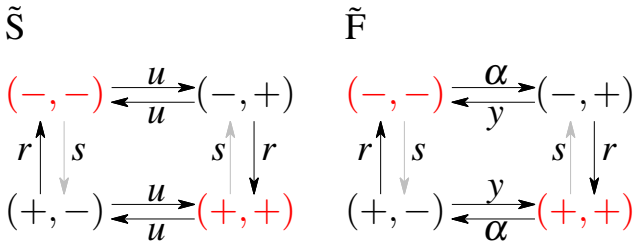

Figure 6. A graphical representation of the optimal circuits without ($\tilde{S}$) and with ($\tilde{F}$) feedback for delayed information transmission with optimized non-steady state initial conditions with a constraint on steady state dissipation $\hat{\sigma}^{ss}$. The exact rate values depend on the value of $\hat{\sigma}^{ss}$ and examples are shown in Figure A1 (model $\tilde{S}$) and Figure A2 (model $\tilde{F}$). The depicted circuits are close to equilibrium. The gray arrow indicates a smaller rate than the black arrow. Optimal non-steady state initial states that have highest probability are shown in red.

To summarize, for all $\hat{\sigma}^{ss} < \infty$, as well as for circuits that have no constraints on $\hat{\sigma}^{ss}$, we found $I(S) < I(\tilde{S})$, $I(F) < I(\tilde{F})$, and $I(S) < I(F)$. Also, for all $\hat{\sigma}^{ss} < \infty$, $I(\tilde{S}) < I(\tilde{F})$, with $I(\tilde{S}) \xrightarrow[\hat{\sigma}^{ss} \to \infty]{} I(\tilde{F})$, where we have defined the optimal mutual information $I(M)$ of a model $M \in \{S, \tilde{S}, F, \tilde{F}\}$.

6.3. Cost of Optimal Information

The maximum information is obtained for maximum *allowed* steady state dissipation. Interestingly the steady state dissipation $\hat{\sigma}^{\mathrm{ss}}$ combined with the circuit topology impose a constraint on the maximum allowed $\Sigma^{\mathrm{avg}}(\tau_p)$. This result follows from the fact that the system strongly relies on the initial condition to increase the information transmitted at small times. Larger μ_0 values allow the system to transmit more information, since the equilibration time is longer. However, fixing the value of $\hat{\sigma}^{\mathrm{ss}}$ constrains the allowed value of μ_0 that determine the initial condition. To gain intuition, additionally to fixing $\hat{\sigma}^{\mathrm{ss}}$, we will fix the mean dissipation $\Sigma^{\mathrm{avg}}(\tau_p)$ until a reset time $\tau_p > \tau$ and find the transition rates returning the optimal mutual information for a chosen readout time $\tau \leq \tau_p$. The results of this optimization presented in Figure 7, show that as Σ^{avg} increases, μ_0 tends towards 1, which corresponds to a probability distribution where only the asymmetric states ($p_0 = (0.5, 0, 0, 0.5)$) are occupied and the transmitted information increases. Further increasing dissipation shows that the $\hat{\sigma}^{\mathrm{ss}}$ constraint can be satisfied in two ways: either by a positive or negative μ_0. Not only does the positive μ_0 transmit more information but the negative μ_0 is forbidden by our choice of $r = 1$. Above a certain value of $\hat{\sigma}^{\mathrm{ss}}$ only the forbidden negative $\mu_0 = -1$ branch corresponding to an initial distribution with all the weight in the anti-aligned states $p_0 = (0, 0.5, 0.5, 0)$ remains (if we chose the counter clockwise solutions by fixing $s = 1$, this probability vector would have been the maximally informative initial state). The system cannot fulfill the constraint of such high dissipation. If we do not constrain $\hat{\sigma}^{\mathrm{ss}}$ we find that the maximum information corresponds to $\mu_0 = 1$ [65], which we report in our analysis below.

We have seen that for both models, if we can choose the initial distribution instead of starting from the steady state, we can significantly increase the transmitted information. What is the "cost" of this choice of initial distribution? To estimate this total cost we calculate the average dissipation during time $\tau_p > \tau$, $\tau_p \Sigma^{\mathrm{avg}}(\tau_p)$, for the circuit with the highest mutual information attainable for a given steady state dissipation rate rate $\hat{\sigma}^{\mathrm{ss}}$ if we allow the initial condition to be out of the steady state (Figure 2). We also introduce the *relaxation cost*, $\tau_p(\Sigma^{\mathrm{avg}} - \hat{\sigma}^{\mathrm{ss}})$ (Figure 8A), as the additional energy dissipated above the steady state value. As argued already, the systems that starts at steady state, i.e., for which $\mu_0 = \mu_0^{\mathrm{ss}}$, will not pay an additional cost (see Figure 2, for $\mu_0 = \mu_0^{\mathrm{ss}}$ the function of $\hat{\sigma}(\tau_p)$ is constant, equal to $\hat{\sigma}^{\mathrm{ss}}$). In this case the mean total dissipation, $\Sigma^{\mathrm{avg}}(\tau_p)$, will be equal to $\hat{\sigma}^{\mathrm{ss}}$ and the relaxation cost goes to zero.

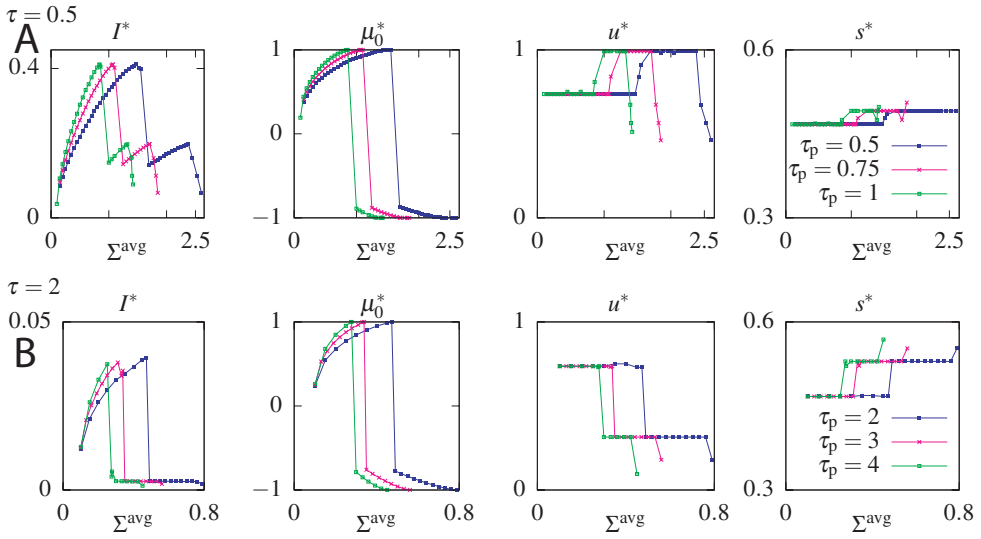

Figure 7. Optimal mutual information (I^*) and optimal parameters μ_0, u, and s for the $\tilde{S}$ model without feedback as function of the average dissipation, Σ^{avg}, for two values of the readout time, $\tau = 0.5$ ((**A**) panels), and $\tau = 2$ ((**B**) panels), and three values of the reset time, τ_p (different colours of curves). Steady state dissipation, $\hat{\sigma}^{\mathrm{ss}}$, was fixed to 0.1.

As shown in Figure 8B, the total cost (z-axis, in colour) generated was only slightly larger for $\tilde{S}$ than for S and the difference is more pronounced only for relatively small $\hat{\sigma}^{ss}$, where the cost in the steady state circuits goes to zero. This result holds for different combinations of delay readout times τ and reset times τ_p, although the value of the total cost naturally increases with τ_p. As discussed above, more information can be transmitted at shorter times and by optimizing over the initial condition.

In order to quantify the intuition that $\tilde{S}$ transmits more information than S at a small price, we plotted in Figure 8C the information gain, $I^* - I^{ss}$, and the relaxation cost with respect to $\tau_p(\hat{\sigma}^{ss})$. $I^* - I^{ss}$ is the difference between the optimal information when the initial distribution is free to be optimized over ($\tilde{S}$) and the optimal information for the system with a steady state initial distribution (S). It quantifies the additional cost from optimizing the initial condition of the gain in information transmission. The relaxation cost is almost the same regardless of the reset time, τ_p. The relaxation cost and the information gain decrease with increasing steady state dissipation, $\hat{\sigma}^{ss}$, as in this regime even the steady state system is able to have slow decorrelation by tuning the switching rates.

This analysis shows that higher optimal mutual information obtained by optimizing over the initial distribution does not generate significantly higher costs. The same result holds when comparing models with feedback F and $\tilde{F}$ (Figure 8D). The information increase from feedback in the $\tilde{F}$ model with optimized initial conditions compared to the F steady state model is minimal at large $\hat{\sigma}^{ss}$ (as expected from Figure 5). While the $\tilde{F}$ model with feedback always transmits more information than the $\tilde{S}$ model without feedback, the total average cost for all $\hat{\sigma}^{ss}$ is smaller for the $\tilde{F}$ model with feedback than for the $\tilde{S}$ model without feedback. This results means that even when feedback does not increase the transmitted information compared to models without feedback, it decreases the total cost.

The information gain of circuits with optimized initial conditions compared to steady state circuits is larger for the $\tilde{S}$ model without feedback than the $\tilde{F}$ model with feedback (Figure 8E) and the relaxation cost decreases monotonically with increasing $\hat{\sigma}^{ss}$. In both the case with and without feedback there is a non-zero and non-infinite value of steady state dissipation where the information gain from optimizing the initial condition is largest. In summary, optimizing the initial condition nearly always incurs a cost, however it absolutely always results in a significant information gain. Table 1 summarizes the comparison of the optimal transmitted information $I(M)$ and the total cost $C(M)$ for all four models $M \in \{S, \tilde{S}, F, \tilde{F}\}$.

Table 1. Comparison between the four models, S, F, $\tilde{S}$, and $\tilde{F}$ in terms of optimal mutual information, I^{opt}, and the cost (value of Σ^{avg} calculated with optimal rates), C.

	I^{opt}	Cost
S, F	$I(S) < I(F)$	$C(S) = C(F)$
$\tilde{S}, \tilde{F}$	$I(\tilde{S}) \leq I(\tilde{F})$	$C(\tilde{S}) > C(\tilde{F})$

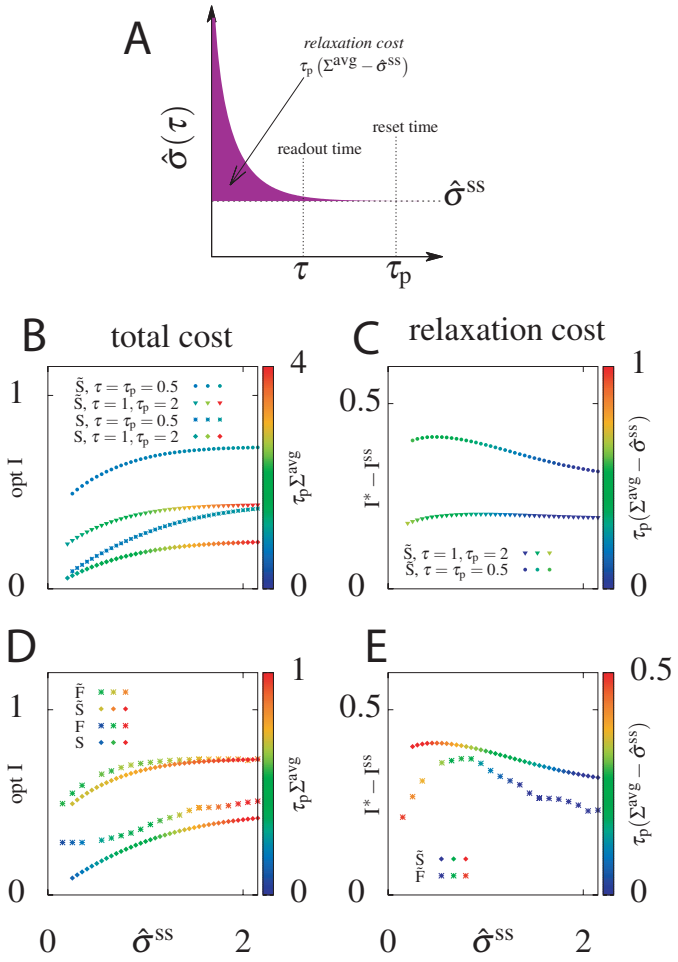

Figure 8. (**A**) Cartoon depicting the relaxation cost (pink area) $\tau_p(\Sigma^{\rm avg} - \hat{\sigma}^{\rm ss})$ of the system equilibrating from a non-steady state initial state, and thus $\hat{\sigma}(\tau) \neq \hat{\sigma}^{\rm ss}$. (**B**) The total cost, $\tau_p\Sigma^{\rm avg}$, of the optimal information transmitted as a function of the steady state entropy dissipation rate, $\tau_p\hat{\sigma}^{\rm ss}$, for models without feedback, that start with the steady state distribution, S, and that optimize the initial distribution, $\tilde{S}$. Results shown for two choices of reset τ_p and readout τ timescales. For the steady state models $\tau_p\Sigma^{\rm avg} = \tau_p\hat{\sigma}^{\rm ss}$. (**C**) The information gain, $I^* - I^{\rm ss}$, of the optimized initital condition model ($\tilde{S}$) compared to the steady state initial condition model (S) and the relaxation cost, $\tau_p(\Sigma^{{\rm avg}-\hat{\sigma}^{\rm ss}})$, as a function of the steady state entropy dissipation rate for the same choices of τ_p and τ as in panel (**B**). (**D**) Comparison of the optimal delayed information and total dissipative cost as a function of the steady state entropy dissipation rate for all four models: without feedback (S, $\tilde{S}$) and with feedback (F, $\tilde{F}$), with the initial distribution equal to the steady state one (S, F) or optimized over ($\tilde{S}$, $\tilde{F}$). $\tau = \tau_p = 0.5$. (**E**) The information gain and relaxation cost of circuits with optimized initial conditions compared to steady state ones for the models with ($\tilde{F}$) and without feedback ($\tilde{S}$). $\tau = \tau_p = 0.5$.

6.4. Suboptimal Circuits

We found the parameters of the stochastic processes, including the initial conditions, that optimally transmit delayed information between the two variables given a constraint on $\hat{\sigma}^{\rm ss}$. However the real initial stimulus may deviate from the optimal one, due to random fluctuations of the environment. To see how an much information an optimal circuit can transmit for different initial conditions, we took the optimal parameters for different fixed $\hat{\sigma}^{\rm ss}$ and readout delay τ, varied the initial condition μ_0 and evaluated the transmitted information and the mean dissipation $\Sigma^{\rm avg}(\tau_p)$ for both models: $\tilde{S}$ and $\tilde{F}$(Figure 9). We find that while information always decreases (Figure 9A,C,E for model $\tilde{S}$ and Figure 9G,I,K for model $\tilde{F}$), as expected, the mean dissipation can be smaller for unexpected values of

the initial condition (Figure 9B,D,F for model $\tilde{S}$ and Figure 9H,J,L for model $\tilde{F}$). The transmitted information of the suboptimal circuits is larger than that of the optimal steady state circuit for many values of μ_0, especially those close to the optimum of the non-steady state circuit ($\mu_0 = 1$). The same conclusions hold for suboptimal circuits with and without feedback. The range of μ_0 values where suboptimal circuits provide an information gain is smaller for circuits with feedback than without feedback, due to the already large information transmission capacity of steady state circuits with feedback.

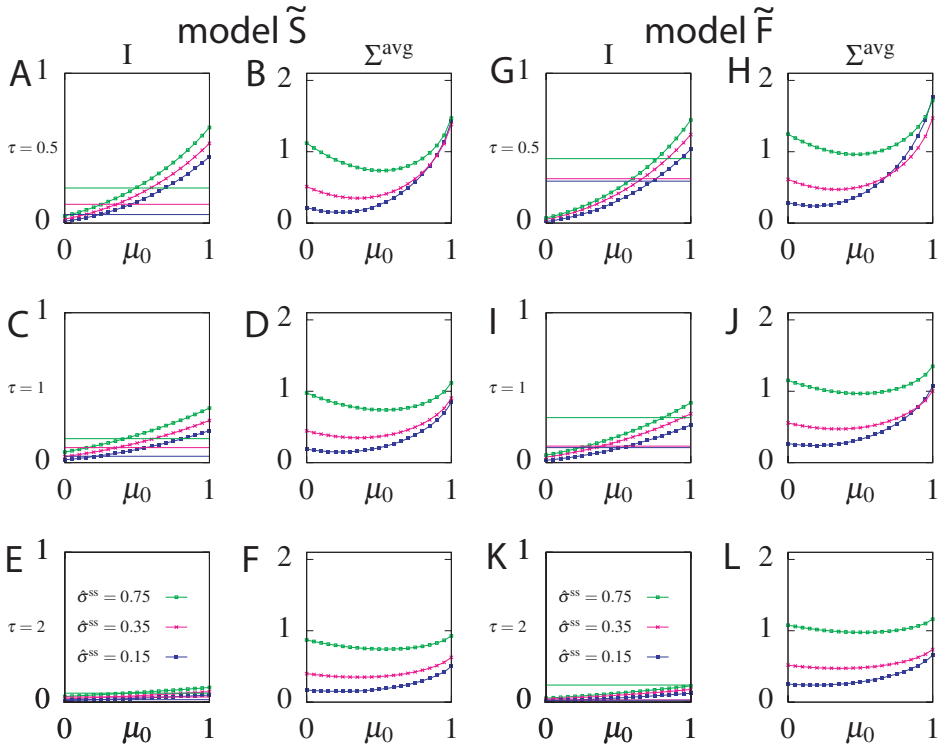

Figure 9. Information for model $\tilde{S}$ (panels (**A**,**C**,**E**)) and model $\tilde{F}$ (panels (**G**,**I**,**K**)) and $\Sigma^{avg}(\tau_p)$ for model $\tilde{S}$(panels (**B**,**D**,**F**)) and model $\tilde{F}$ (panels (**H**,**J**,**L**)) of information-optimal circuits with $\mu_0 = 1$ evaluated for different values of the initial condition μ_0. The circuits parameters are evaluated by optimizing information transmission for $\tau = 0.5$ (**A**,**B**), $\tau = 1$ (**C**,**D**) and $\tau = 2$ (**E**,**F**) and fixed $\hat{\sigma}^{ss} = 0.15$ (blue lines), $\hat{\sigma}^{ss} = 0.35$ (magenta lines), $\hat{\sigma}^{ss} = 0.75$ (green lines). $\tau_p = \tau$ in all plots. For comparison we plot the optimal information of the steady state circuit S and F, respectively, optimized for the same steady state dissipation $\hat{\sigma}^{ss}$ and readout delay τ (solid lines). The information always decreases for non-optimal values of μ_0 but the mean dissipation can be smaller for unexpected initial conditions.

7. Gene Regulatory Circuits

The coupled two state system model considered above can be thought of as a simplified model of receptor—ligand binding. It can also be considered as an overly simplified model of gene regulation where the input variable describes the presence or absence of a transcription factor and the output—the activation state of the regulated gene. However, the continuous nature of transcription factor concentrations has proven important when considering information transmission in these systems [53,54]. We will not repeat the whole optimization problem for continuous variables but we calculate and discuss the form of dissipation in the simplest gene regulatory module that can function out of equilibrium.

7.1. Bursty Gene Regulation

The simplest gene regulatory system that can function out of equilibrium is a model that accounts for transcriptional bursts [67–73]. The promoter state has two possible states: a basal expression state where the gene is read out a basal rate R_0 and an activated expression state where the gene is read out at rate R_1. The promoter switches between these two states by binding a transcription factor present at concentration c, with rate k_+ and unbinds at a constant rate k_-. The probability that there are g product proteins of this gene in the cell (we integrate out the mRNA state due to a separation of timescales) is $P(g) = P_0(g) + P_1(g)$, where $P_0(g)$ describes the probability that the promoter is in the basal state and there are g proteins and $P_1(g)$ describes the analogous probability for the promoter to be in the activated state. The probability distribution evolves both due to binding and unbinding of the transcription factor and to protein production and degradation (with rate τ^{-1}) according to

$$\begin{aligned}\frac{dP_0(g)}{dt} &= \frac{g+1}{\tau}P_0(g+1) + k_-P_1(g) + R_0P_0(g-1) + \\ &\quad - \left(k_+c + \frac{g}{\tau} + R_0\right)P_0(g), \end{aligned} \tag{20}$$

$$\begin{aligned}\frac{dP_1(g)}{dt} &= \frac{g+1}{\tau}P_1(g+1) + k_+cP_0(g) + R_1P_1(g-1) + \\ &\quad - \left(k_- + \frac{g}{\tau} + R_1\right)P_1(g). \end{aligned} \tag{21}$$

These equations can be solved analytically in steady state in terms of special functions [79,80]. In the limit of fast promoter switching (k_+ and k_- go to infinity and their ratio $K \equiv k_+/k_-$ is constant) the system is well described by a Poisson distribution

$$P_1^*(g) = \frac{1}{1+cK}\frac{(R_{ef}\tau)^g}{g!}\exp(-R_{ef}\tau) = cKP_0^*(g) \tag{22}$$

where R_{eff} is an effective production rate:

$$R_{eff} = \frac{k_+cR_1 + k_-R_0}{k_+c + k_-}. \tag{23}$$

The total steady state dissipation $\sigma^{ss} = \sigma_0 + \sigma_1 + \sigma_2$ calculated from Equation (11) can be split into three parts, where

$$\sigma_0 = \sum_g \left(P_0^*(g)k_+c - P_1^*(g)k_-\right)\log\frac{k_+c}{k_-}, \tag{24}$$

$$\sigma_1 = \sum_g \left(P_0^*(g)R_0\log(R_0\tau) + P_1^*(g)R_1\log(R_1\tau)\right), \tag{25}$$

$$\begin{aligned}\sigma_2 &= -\sum_g P_0^*(g)\left[R_0\log(g+1) + \frac{g}{\tau}\log\frac{R_0\tau}{g}\right] + \\ &\quad -\sum_g P_1^*(g)\left[R_1\log(g+1) + \frac{g}{\tau}\log\frac{R_1\tau}{g}\right]. \end{aligned} \tag{26}$$

The first two expressions can be simplified using the normalization relations $\sum_g \left(P_0^*(g) + P_1^*(g)\right) = 1$ and $\sum_g P_1^*(g) = \frac{k_+c}{k_- + k_+c}$ obtaining:

$$\sigma_0 = 0 \tag{27}$$

$$\sigma_1 = \frac{1}{k_- + k_+c}\left(R_0\log(R_0\tau)k_- + R_1\log(R_1\tau)k_+c\right). \tag{28}$$

We now use these results to examine steady state dissipation in the equilibrium limit and the limit of the fast switching promoter. Similar results but in slightly different limits were obtained in reference [30].

Equilibrium Limit. Equilibrium is surely achieved if there is only one promoter state. In terms of our model this corresponds to k_+ is vanishing and $k_- \neq 0$. In this limit the activated state is never occupied and the steady state probability goes to $P_1^*(g) \equiv 0$. Equations (21) and (22) result in a Poisson distribution with mean $R_0\tau$ and we can verify that detailed balance is satisfied

$$P_0^*(g)W_{(g\to g\pm 1)} = P_0^*(g\pm 1)W_{(g\pm 1\to g)}, \tag{29}$$

as confirmed by $\sigma_2 = -\sigma_1$ in Equations (25)–(27).

Fast promoter switching limit. In the fast promoter switching limit the dissipation of the system is:

$$\sigma_{FS} = \frac{cK}{(1+cK)^2}(R_0 - R_1)\mathrm{Log}\left(\frac{R_0}{R_1}\right). \tag{30}$$

σ_{FS} is always positive, but the equilibrium regime is reached only if k_- or k_+ asymptotically vanish. For finite binding and unbinding rates the system is not in equilibrium despite being well described by an equilibrium-like steady state probability distribution. Since this example is mainly presented as a pedagogical application of dissipation, for completeness we derive similar results in the Langevin description in Appendix D, discussing the differences in dissipation arising from model coarse graining [81–83].

8. Discussion

All living organisms, even the most simple ones, in order to adapt to the environment, must read and process information. In the case of cells, transmitting information means sensing chemical stimuli via receptors and activating biochemical pathways in response to these signals. Such reading and transmitting signals comes at a price—it consumes energy. There are plenty of possible designs of these regulatory circuits, yet not all of them are found in nature [2]. The question arises why some network regulatory functions are frequent and others non-existing. One way to approach such a question is to optimize a (specific) function by choosing the circuit's regulatory function. The choices of optimzed function that have been considered include noise (minimization) [11], time-delay of response (minimization) [2] or information transmitted between the input and output (maximization) [56].

Two different circuits can produce and use the same amount of proteins, but the energy dissipated in them is different. In other words, we assume that while ATP is certainly needed in a molecular circuit, it is part of the hardware of the network and cannot be modified a lot. Instead, we asked about the best regulatory functions (software) we can implement, given a certain set of hardware. For this reason we worked with a simplified binary representation of the circuits to concentrate on the regulatory computation and turned the problem of finding the optimal regulatory function into finding the optimal parameters of stochastic processes.

Our main previous findings about steady state circuits can be related to tasks performed by the circuits [66]. Circuits that function close to equilibrium transmit information optimally using positive feedback loops that are characteristic of long-term readouts responsible for cell fate commitment [84,85]. Circuits that function far from equilibrium transmit information using negative feedback loops that are representative of shock responses that are transient but need to be fast [86,87]. Therefore cells may implement non-equilibrium solutions when fast responses are needed and rely on equilibrium responses when averaging is possible and there is no rush. This results agrees with the general finding of Lan et al. [29] for continuous biochemical kinetics that negative feedback circuits always break detailed balance and such circuits function out of equilibrium.

In general in steady state we find that models with feedback significantly outperform models without feedback in terms of optimal information transmission between the two variables, but the respective costs of optimal information transmission are the same. Circuits close and far to equilibrium

rely on a mixture of slow and fast timescales to delay relaxation and transmit information. The only other solution available in our simple setting is using the initial condition, which is efficient in terms of information transmission but costly.

Here we identified two properties linked to feedback: it does not necessarily transmit more information if we are allowed to pick an optimal initial condition compared to a system without feedback. Yet in this case implementing a circuit with feedback can reduce the non-equilibrium costs. In general, introducing an optimized intitial condition incurs a cost, but this cost is often minimal, especially taking into account the information gained. This cost is interpretable biologically as the external energetic cost needed to place the system in a specific initial condition. This cost must be provided by the work of another regulatory element or circuit or an external agent or force. This specific initial condition requires poising the system in a specific point. Yet it does not seem biologically implausible, let alone impossible, to "prepare" the intitial state after cell division or mitosis, or upon entering a new phase of the cell cycle [88]. For example, a specific gene expression state or receptor state (e.g., $(+,+)$ or $(-,-)$) seems easily attainable. Modifying the initial conditions from the optimal μ_0 in circuits that function out of steady state decreases the transmitted information but can also decrease the mean dissipation. Therefore optimizing preparing the system out of steady state may still be a useful strategy for transmitting information.

One could look at these results from two perspectives: on the one hand argue that circuits with feedback transmit more information in the steady state setting; on the other hand feedback exhibits frugality in expenses in the case of optimized initial distributions. One could also defend the models without feedback stating that they can be only slightly worse in terms of information transmission (optimized initial distribution case) and can be found to dissipate the same amount of energy (steady state initial distribution). All circuits will reach steady state, however especially during fast processes such as development [89] or stress response [87], the information transmitted during short times may be what matters for downstream processes. In general regardless of the timescale, circuits with feedback perform better (or equally well) than regulatory system with no feedback, both in terms of information transmission and the cost of transmitting this optimal information.

The learning rate is another quantity that has been useful in studying bipartite systems in stochastic thermodynamics [43–45]. The learning rate, defined as $l_x = \partial_\tau I[z_\tau, x_{t+\tau}]|_{\tau=0}$, gives the instantaneous increase in information that the output variable has by continuing to learn about the input variable. We calculate the learning rate for our informationally-optimal models when they are in steady state (Figure A3). For models without feedback the learning rate is bounded by σ_x (as defined in Appendix E), such that $\eta = \ell_x/\sigma_x \leq 1$. It this case the interpretation of the learning rate allows us to estimate how closely the output variable is following the input variable and positive learning rates are indicative of adaptation and learning. Not surprisingly we find that the model with steady state initial conditions has a larger learning rate than the model with optimized initial conditions since model $\tilde{A}$ relies less on the parameters of the network than model A to transmit information and more on the initial conditions (that are forgotten in the steady state calculation). Calculating a time delay dependent learning rate would be more informative. The learning rate also increases with $\hat{\sigma}$, in agreement with previous statements that learning is easier far from equilibrium [29,30,43]. We also performed the same calculation for models with feedback but as was pointed out previously [44,90,91], the interpretation of the learning rate becomes less clear in these systems since input and output are no longer clearly defined. Instead the above one-sided definition should be replaced by a time integral over the trajectory to distinguish if the learning is of the other variable (z) or a previous instance of the same variable ($x_{t-\tau}$). The calculated quantity instead tells us about the ability of x to respond to z, assuming z was fluctuating freely. In that sense a positive value of l_x tells us that the dynamics of the two variables of the circuit are not completely decoupled in steady state, except in the case of model F close to equilibrium. Our results tell us that equilibrium imposes a symmetry between input and output, which is broken either by initial conditions ($\tilde{F}$ at small $\hat{\sigma}$) or large dissipation.

Lastly, for pedagogical purposes we attempted to discuss the link between dissipation calculations that are often performed on binary regulatory systems and continuous variables, showing that the simplest model of bursty transcription can result in non-zero dissipation, even in the fast switching limit where the steady state equilibrium Poisson distribution is recovered. Bursty gene expression is wide spread from bacteria [71,72], yeast [92] to invertebrates [73,89] and mammals [68]. Bursty self-activating genes in intermediate fast switching regimes have also been shown to have different stability properties than pure equilibrium systems, due to non-equilibrium cycling through the coupled promoter and protein states [93]. While cells are not energy limited, the discussion recounted in this paper may suggest that different modes of regulation (including burstiness) may be better suited to slow and fast responses.

Author Contributions: Conceptualization, P.S.-R., D.V. and A.M.W.; Formal analysis, P.S.-R., D.V., J.M. and A.M.W.; Funding acquisition, A.M.W.; Investigation, P.S.-R., D.V., J.M. and A.M.W.; Methodology, P.S.-R., D.V. and A.M.W.; Project administration, J.M. and A.M.W.; Software, P.S.-R. and D.V.; Supervision, J.M. and A.M.W.; Visualization, P.S.-R. and A.M.W.; Writing—original draft, P.S.-R., D.V. and A.M.W.; Writing—review & editing, P.S.-R., D.V., J.M. and A.M.W.

Funding: This research was funded by FP7 People: Marie-Curie Actions grant number 303561 and Narodowe Centrum Nauki grant number 2015/17/B/ ST1/00693.

Acknowledgments: We thank T. Lipniacki, A. Nourmohammad and T. Mora for helpful discussions. This work was in part supported by MCCIG no. 303561. J. M. would like to thank the National Science Centre (Poland) for financial support under Grant No. 2015/17/B/ ST1/00693.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Model without Feedback

The transition matrix for the model without feedback reads:

$$\mathcal{L} = \begin{pmatrix} u+s & -u & -r & 0 \\ -u & u+r & 0 & -s \\ -s & 0 & u+r & -u \\ 0 & -r & -u & u+s \end{pmatrix}, \tag{A1}$$

where the rates are defined in Figure 1A. By matrix diagonalization we find the eigenvalues and eigenvectors and calculate the probability distribution $p(x_\tau, z_0)$ at time τ for the four states,

$$\begin{aligned} &p_{+,+}(\tau) = p_{-,-}(\tau) = \\ &\frac{e^{-\frac{\tau(s+2u+1)}{\lambda}}}{4(s-2u+1)}\Big((\mu_0(s-2u+1)+s-1)e^{\frac{2\tau u}{\lambda}}+ \\ &(s-1)\left(-e^{\frac{(s+1)\tau}{\lambda}}\right)+(s-2u+1)e^{\frac{\tau(s+2u+1)}{\lambda}}\Big), \end{aligned} \tag{A2}$$

and

$$\begin{aligned} &p_{+,-}(\tau) = p_{-,+}(\tau) = \\ &\frac{e^{-\frac{\tau(s+2u+1)}{\lambda}}}{4(s-2u+1)}\Big(-(\mu_0(s-2u+1)+s-1)e^{\frac{2\tau u}{\lambda}}+ \\ &(s-1)e^{\frac{(s+1)\tau}{\lambda}}+(s-2u+1)e^{\frac{\tau(s+2u+1)}{\lambda}}\Big). \end{aligned} \tag{A3}$$

The steady state distribution is given by the eigenvector corresponding to the zeroth eigenvalue,

$$p^\infty = \left(\frac{u+1}{2s+4u+2'}, \frac{s+u}{2s+4u+2'}, \frac{s+u}{2s+4u+2'}, \frac{u+1}{2s+4u+2}\right). \tag{A4}$$

These results allow us to calculate

$$\mu = \frac{-e^{-2ut}(1-s)}{1+s-2u} + \frac{-e^{-(1+s)t}(\mu_0(1+s-2u)-(1-s))}{1+s-2u} \tag{A5}$$

$$= \mu_0 e^{-(1+s)t} + \frac{1-s}{1+s-2u}\left(e^{-(1+s)t} - e^{-(1+s)t}\right). \tag{A6}$$

Appendix B. Model with Feedback

The transition matrix for the model with feedback reads defined in Figure 1B:

$$\mathcal{L} = \begin{pmatrix} \alpha+s & -y & -r & 0 \\ -\alpha & y+r & 0 & -s \\ -s & 0 & y+r & -\alpha \\ 0 & -r & -y & s+\alpha \end{pmatrix}. \tag{A7}$$

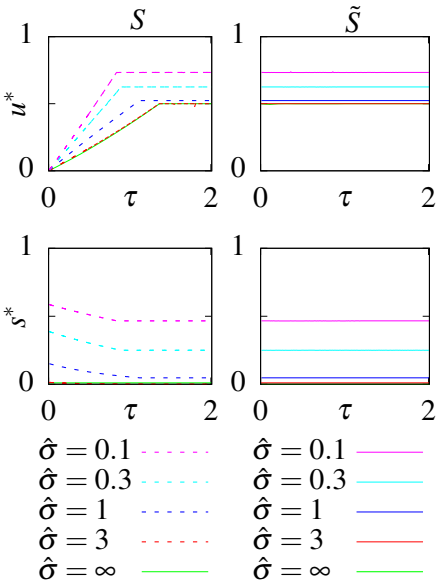

Figure A1. The optimal parameters as a function of the readout delay, τ, for the models without feedback, S and $\tilde{S}$, at different constrained steady state dissipation rates $\hat{\sigma}^{ss}$.

The detailed derivation of the steady state quantities and eigenvalues is given in Mancini et al. [66]. Here we just summarize the main results. The steady state probability distribution is:

$$p^{\infty} = \frac{1}{2A}\{1+y, s+\alpha, s+\alpha, 1+y\}, \tag{A8}$$

where we have defined A and ρ in Equations (18) and (19).

The eigenvalues of the matrix in Equation (A7) are $\{\lambda_i\} = \{0, A, (A-\rho)/2, (A+\rho)/2\}$ and $\lambda = (A-\rho)/2$ is always the smallest eigenvalue. For a model with steady state initial conditions μ reads

$$\mu = \exp\left(-\frac{A}{2\lambda}\tau\right)\left\{q\cosh\left(\frac{\rho}{2\lambda}\tau\right) - \frac{\left[s^2-(1+y)^2-4\alpha+\alpha^2+2s(2y+\alpha)\right]}{A\rho}\sinh\left(\frac{\rho}{2\lambda}\tau\right)\right\}, \tag{A9}$$

with $q = (1+y-s-\alpha)/A$ and the rescaled time $\tau = t\lambda$.

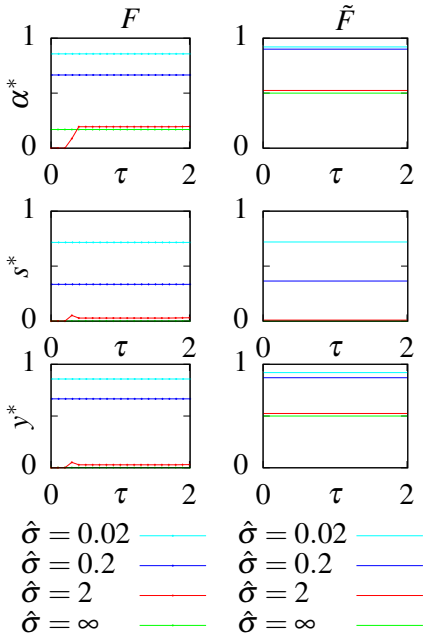

Figure A2. The optimal parameters as a function of the readout delay τ for models with feedback, F and $\tilde{F}$, at different constrained steady state dissipation rates $\hat{\sigma}^{\mathrm{ss}}$.

Appendix C. Entropy Production Rate

In this Appendix we present an alternative derivation of dissipation. We denote probability of state i by p_i and the entropy of the distribution is defined as:

$$S(t) = -\sum_i p_i(t) \log p_i(t). \tag{A10}$$

The entropy production rate formula is derived by differentiating the entropy with respect to time:

$$\begin{aligned} \dot{S}(t) &= -\sum_i \dot{p}_i(t) \log p_i(t) - \sum_i p_i(t) \frac{1}{p_i(t)} \dot{p}_i(t) \\ &= -\sum_i \dot{p}_i(t) \log p_i(t) - \left(\sum_i p_i(t) \right)'. \end{aligned}$$

Denoting by w_{ij} the transition rate from state i to state j, we obtain $\dot{p}_i(t) = \sum_{j \neq i} w_{ji} p_j(t) - w_{ij} p_i(t)$. We define w_{ii} as $-\sum_{j, j \neq i} w_{ij}$, so that we can write compactly $\dot{p}_i(t) = \sum_j p_j(t) w_{ji}$ and the expression for $\dot{S}(t)$ becomes:

$$\begin{aligned} \dot{S}(t) &= -\sum_i \left(\sum_j w_{ji} p_j(t) \right) \log p_i(t) - 0 \\ &= -\sum_{i,j} w_{ji} p_j(t) \log p_i(t). \end{aligned} \tag{A11}$$

With the definition of w_{ii}, the terms w_{ij} satisfy $\sum_j w_{ij} = 0$. The following expression $-\sum_i p_i(t) \log p_i(t) \sum_j w_{ij} = -\sum_{i,j} w_{ij} \log p_i(t)$ is then equal to zero and we subtract it form (A11) to obtain a compact form:

$$\begin{aligned} \dot{S}(t) &= \left(\sum_{i,j} p_i(t) w_{ij} \log p_i(t) - \sum_{i,j} p_i(t) w_{ij} \log p_j(t) \right) \\ &= \sum_{i,j} p_i(t) w_{ij} \log \tfrac{p_i(t)}{p_j(t)}. \end{aligned} \tag{A12}$$

Further formula manipulation gives:

$$\begin{aligned} \dot{S}(t) &= \frac{1}{2} \sum_{i,j} p_i(t) w_{ij} \log \frac{p_i(t)}{p_j(t)} + \frac{1}{2} \sum_{j,i} p_j(t) w_{ji} \log \frac{p_j(t)}{p_i(t)} \\ &= \frac{1}{2} \sum_{i,j} p_i(t) w_{ij} \log \frac{p_i(t)}{p_j(t)} - \frac{1}{2} \sum_{j,i} p_j(t) w_{ji} \log \frac{p_i(t)}{p_j(t)} \\ &= \frac{1}{2} \sum_{i,j} \left(p_i(t) w_{ij} - p_j(t) w_{ji} \right) \log \frac{p_i(t)}{p_j(t)} \\ &= \underbrace{\frac{1}{2} \sum_{i,j} \left(p_i(t) w_{ij} - p_j(t) w_{ji} \right) \log \frac{w_{ji}}{w_{ij}}}_{\textit{entropy flow}} + \\ &\underbrace{\frac{1}{2} \sum_{i,j} \left(p_i(t) w_{ij} - p_j(t) w_{ji} \right) \log \frac{p_i(t) w_{ij}}{p_j(t) w_{ji}}}_{\textit{entropy production rate}}. \end{aligned} \tag{A13}$$

The difference between the entropy production rate and the entropy flow, is the rate at which the whole entropy of a system changes. The entropy flow quantifies the flux of entropy from the system to the outside. In the steady state, as the entropy does not change, the two terms are equal, which means that the whole entropy produced by the system is dissipated.

The second underbracket of (Equation (A13)) can be rewritten in the familiar form:

$$\sigma(t) = \sum_{i,j} p_i(t) w_{ij} \log \frac{p_i(t) w_{ij}}{p_j(t) w_{ji}}. \tag{A14}$$

Appendix D. Langevin Description of Bursty Gene Regulation

A bursty model of transcription such as the one presented in Section 7.1 can be written in a Langevin description introducing the frequency for the promoter to be in the activated state n [53]:

$$\frac{dn}{dt} = -ck_+ n - k_- n + \xi_n, \tag{A15}$$

$$\frac{dg}{dt} = Rn - \frac{1}{\tau} g + \xi_g, \tag{A16}$$

where the fluctuations are given by

$$\langle \xi_n(t) \xi_n(t') \rangle = 2(k_+ c(1 - \bar{n}) + k_- \bar{n}) \delta(t - t'), \tag{A17}$$

$$\langle \xi_g(t) \xi_g(t') \rangle = 2(R\bar{n} + \bar{g}/\tau) \delta(t - t'). \tag{A18}$$

These equations describe the fluctuations of the promoter state and the protein concentration g around the equilibrium solution $(\bar{n}, \bar{g}) = \left(\frac{k_+ c}{k_- + k_+ c}, \frac{k_+ cR\tau}{k_- + k_+ c} \right)$. In order to lighten notations, we have

used (n, g) instead of the standard form $(\delta n, \delta g)$ to describe fluctuations. Equations (A15) and (A16) can be recast into the matrix form form $\dot{\mathbf{X}} = -A\mathbf{X} + \boldsymbol{\xi}$ with

$$A = \begin{pmatrix} ck_+ + k_- & 0 \\ -R & \frac{1}{\tau} \end{pmatrix}, \tag{A19}$$

and the noise correlation matrix $\langle \boldsymbol{\xi}(t)\boldsymbol{\xi}(t')\rangle = 2D\delta(t-t')$ is

$$D = \begin{pmatrix} k_+ c(1-\bar{n}) + k_- \bar{n} & 0 \\ 0 & (R\bar{n} + \bar{g}/\tau) \end{pmatrix}. \tag{A20}$$

The correlation matrix Σ can be computed with standard methods [56], by inverting the relation $D = A\Sigma + \Sigma A^t$:

$$\begin{aligned} \Sigma \;=\; & \begin{pmatrix} \langle nn\rangle & \langle ng\rangle \\ \langle gn\rangle & \langle gg\rangle \end{pmatrix} = \frac{1}{(ck_+ + k_-)^2} \cdot \\ & \begin{pmatrix} 2ck_-k_+ & \frac{2ck_-k_+R\tau}{((ck_+\tau + k_-\tau + 1)} \\ \frac{2ck_-k_+R\tau}{(ck_+\tau + k_-\tau + 1)} & \frac{2ck_+R\tau\left(\tau\left((ck_+ + k_-)^2 + k_-R\right) + ck_+ + k_-\right)}{(ck_+\tau + k_-\tau + 1)} \end{pmatrix}. \end{aligned} \tag{A21}$$

Entropy Production

The probability of a trajectory of a multivariate Langevin process can be calculated via the Onsager-Machlup formalism. Using this probability as starting point, the dissipation can be exactly derived (see Ref. [94], where the computation is done in detail and in a self-contained fashion). For the case of symmetric variables under time reversal the entropy production can be written in a compact form, where we have the index k runs over all the variables:

$$W(t) = \sum_k D_{kk}^{-1} \int_0^t ds\, (A\mathbf{X})_k \dot{X}_k. \tag{A22}$$

In our case, using Equations (A19) and (A20), one has

$$\begin{aligned} W(t) \;=\; & D_{nn}^{-1} \textstyle\int_0^t dt'\, (ck_+ n(t') - k_- n(t'))\dot{n}(t') + \\ & D_{gg}^{-1} \textstyle\int_0^t dt'\, (Rn(t') - \frac{1}{\tau} g(t'))\dot{g}(t'). \end{aligned} \tag{A23}$$

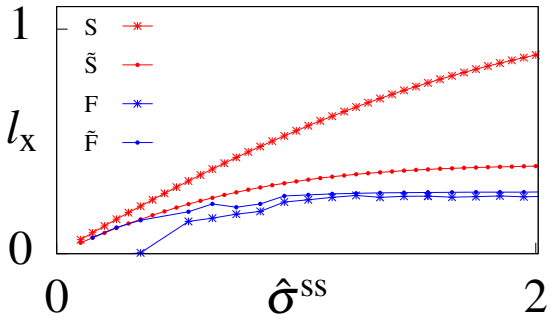

Figure A3. The learning rate for the output variable x as a function of the rescaled steady state dissipation, $\hat{\sigma}^{ss}$, calculated at steady state for models with (F and $\tilde{F}$) and without feedback (S and $\tilde{S}$). Models $\tilde{S}$ and $\tilde{F}$ have optimized initial conditions (that do not enter this calculations except for the optimal parameters) and models S and F are constrained to have initial conditions in steady state.

Equation (A23) can be simplified by considering that all terms which are exact derivatives are not extensive in time (terms like $\int_0^t dt' n(t')\dot{n}(t') = \frac{1}{2}\left(n^2(t) - n^2(0)\right)$ or its equivalent in g can be neglected in the large t limit). All the steady state correlations are also time translational invariant, i.e., $\left\langle \int_0^t dt' n(t')\dot{g}(t') \right\rangle \equiv t \langle n\dot{g} \rangle$. As a result, the dissipation is:

$$\sigma^{LE} = \lim_{t\to\infty} \frac{\langle W_t \rangle}{t} = \frac{R}{R\overline{n} + \frac{\overline{g}}{\tau}} \langle n\dot{g} \rangle . \tag{A24}$$

The correlation $\langle n\dot{g} \rangle$ in Equation (A24) can be computed by replacing $\dot{g}$ with Equation (A16), yielding $\langle n\dot{g} \rangle = R \langle nn \rangle - \frac{1}{\tau} \langle ng \rangle$. Substituting this expression into Equation (A21) we obtain:

$$\sigma^{LE}(c) = \frac{R\tau(1 - ck_+\tau_s)}{\tau + \tau_s}, \tag{A25}$$

where $\tau_s = (ck_+ + k_-)^{-1}$. Note that in the limit $\tau_s \to 0$ the dissipation is not dependent on c and equal to $\sigma_0^{LE} = R/(1 + cK)$, where K is equal to k_+/k_-. Additionally, for $K \to \infty$ (which corresponds to no flux to the inactive state, $k_- \to 0$) the dissipation vanishes like in the master equation formulation (Equation (30)).

As final remark, we note that a Langevin formulation is a coarse grained description of the Master equation approach described in Section 7.1. This kind of coarse graining procedure integrates away degrees of freedom which can carry non-equilibrium currents and can lead to lower values of dissipation [81–83]. For instance, consider the limit $R_0 \approx \epsilon$ small but finite, Equation (30) becomes $\sigma^{ME} = cK(1 + R_1)/(1 + cK)^2 \log \frac{R_1}{\epsilon}$ and one finds $\sigma^{ME} > \sigma^{LE}$.

Appendix E. Learning Rate

Lastly, following Barato et al. [43] we consider the learning rate in steady state (we limit ourselves to the steady state discussion since it allows us to get analytical intuition)

$$l_x = -\sum_i p_i^\infty \sum_{i\neq j} w_{ij} \log \frac{p_i^\infty}{p_j^\infty}, \tag{A26}$$

which was defined to describe the rate at which the output x learns about the dynamics of the stochastic input z. For our system, the learning rate is explicitly given by

$$\begin{aligned} -l_x \quad = \quad & p_1^\infty w_{13} \log \tfrac{p_1^\infty}{p_3^\infty} + p_3^\infty w_{31} \log \tfrac{p_3^\infty}{p_1^\infty} \\ & p_2^\infty w_{24} \log \tfrac{p_2^\infty}{p_4^\infty} + p_4^\infty w_{42} \log \tfrac{p_4^\infty}{p_2^\infty}, \end{aligned} \tag{A27}$$

and is bounded by σ_x defined as:

$$\begin{aligned} \sigma_x \quad = \quad & (p_1^\infty w_{13} - p_3^\infty w_{31}) w_{13} \log \tfrac{w_{13}}{w_{31}} \\ & (p_2^\infty w_{24} - p_4^\infty w_{42}) \log \tfrac{w_{24}}{w_{42}}. \end{aligned} \tag{A28}$$

For the models without feedback (S and $\tilde{S}$) the learning rate is:

$$l_x = \frac{u(s-1)}{s + 2u + 1} \log \frac{u + s}{u + 1}. \tag{A29}$$

In models without feedback $w_{12} = w_{21}$ and $w_{34} = w_{43}$ and the steady state dissipation rate comes only from the output, x ($\sigma_z = 0$ and $\sigma_x = \sigma^{ss}$) and is given by Equation (15), such that

$$\eta = \frac{l_x}{\sigma_x} = \log \frac{u + s}{u + 1} / \log s \leq 1. \tag{A30}$$

For models with feedback (F and $\tilde{F}$) the learning rate is harder to interpret since the input no longer changes independently of the output. Formally we can still calculate the quantity in Equation (A27) as

$$l_x = \frac{y(s-\alpha)}{\alpha+s+y+1} \log \frac{y+1}{\alpha+1}, \tag{A31}$$

and

$$\sigma_x = \frac{y(s-\alpha)}{\alpha+s+y+1} \log s. \tag{A32}$$

The informational efficiency is:

$$\eta = \frac{y+1}{\alpha+1} / \log s, \tag{A33}$$

which is bounded by 1 only if $sy \leq \alpha$ (see Figure A3).

References

1. Bialek, W. *Biophysics*; Princeton University Press: Princeton, NJ, USA, 2012.
2. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*; Chapman & Hall: Boca Raton, FL, USA, 2006.
3. Phillips, R.; Kondev, J.; Theriot, J.; Garcia, H. *Physical Biology of the Cell*; Garland Science: New York, NY, USA, 2012; p. 2012.
4. Hopfield, J. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 4135–4139. [CrossRef] [PubMed]
5. Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **1975**, *57*, 587–595. [CrossRef]
6. McKeithan, T.W. Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5042–5046. [CrossRef]
7. Tostevin, F.; Howard, M. A stochastic model of Min oscillations in Escherichia coli and Min protein segregation during cell division. *Phys. Biol.* **2006**, *3*, 1. [CrossRef] [PubMed]
8. Tostevin, F.; Howard, M. Modeling the Establishment of {PAR} Protein Polarity in the One-Cell C. elegans Embryo. *Biophys. J.* **2008**, *95*, 4512–4522. [CrossRef] [PubMed]
9. François, P.; Hakim, V. Design of genetic networks with specified functions by evolution in silico. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 580–584. [CrossRef]
10. François, P.; Hakim, V.; Siggia, E.D. Deriving structure from evolution: Metazoan segmentation. *Mol. Syst. Biol.* **2007**, *3*, 154. [CrossRef]
11. Saunders, T.E.; Howard, M. Morphogen profiles can be optimized to buffer against noise. *Phys. Rev. E* **2009**, *80*, 041902. [CrossRef] [PubMed]
12. Tkacik, G.; Callan, C.G.; Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12265–12270. [CrossRef]
13. Mehta, P.; Goyal, S.; Long, T.; Bassler, B.L.; Wingreen, N.S. Information processing and signal integration in bacterial quorum sensing. *Mol. Syst. Biol.* **2009**, *5*, 325. [CrossRef]
14. Bintu, L.; Buchler, N.E.; Garcia, H.G.; Gerland, U.; Hwa, T.; Kondev, J.; Phillips, R. Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **2005**, *15*, 116–124. [CrossRef] [PubMed]
15. Bintu, L.; Buchler, N.E.; Garcia, H.G.; Gerland, U.; Hwa, T.; Kondev, J.; Kuhlman, T.; Phillips, R. Transcriptional regulation by the numbers: Applications. *Curr. Opin. Genet. Dev.* **2005**, *15*, 125–135. [CrossRef] [PubMed]
16. Garcia, H.G.; Phillips, R. Quantitative dissection of the simple repression input-output function. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12173–12178. [CrossRef] [PubMed]
17. Kuhlman, T.; Zhang, Z.; Saier, M.H.; Hwa, T. Combinatorial transcriptional control of the lactose operon of Escherichia coli. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6043–6048. [CrossRef]
18. Dubuis, J.O.; Tkacik, G.; Wieschaus, E.F.; Gregor, T.; Bialek, W. Positional information, in bits. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16301–16308. [CrossRef]
19. Tostevin, F.; ten Wolde, P.R. Mutual Information between Input and Output Trajectories of Biochemical Networks. *Phys. Rev. Lett.* **2009**, *102*, 218101. [CrossRef]

20. Tostevin, F.; ten Wolde, P.R. Mutual information in time-varying biochemical systems. *Phys. Rev. E* **2010**, *81*, 061917. [CrossRef]
21. de Ronde, W.H.; Tostevin, F.; ten Wolde, P.R. Effect of feedback on the fidelity of information transmission of time-varying signals. *Phys. Rev. E* **2010**, *82*, 031914. [CrossRef]
22. Savageau, M. Design of molecular control mechanisms and the demand for gene expression. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5647–5651. [CrossRef]
23. Scott, M.; Gunderson, C.W.; Mateescu, E.M.; Zhang, Z.; Hwa, T. Interdependence of cell growth and gene expression: Origins and consequences. *Science* **2010**, *330*, 1099–1102. [CrossRef]
24. Aquino, G.; Tweedy, L.; Heinrich, D.; Endres, R.G. Memory improves precision of cell sensing in fluctuating environments. *Sci. Rep.* **2014**, *4*, 5688. [CrossRef] [PubMed]
25. Vergassola, M.; Villermaux, E.; Shraiman, B.I. 'Infotaxis' as a strategy for searching without gradients. *Nature* **2007**, *445*, 406–409. [CrossRef] [PubMed]
26. Celani, A.; Vergassola, M. Bacterial strategies for chemotaxis response. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1391–1396. [CrossRef] [PubMed]
27. Siggia, E.D.; Vergassola, M. Decisions on the fly in cellular sensory systems. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E3704–E3712. [CrossRef]
28. Cheong, R.; Rhee, A.; Wang, C.J.; Nemenman, I.; Levchenko, A. Information Transduction Capacity of Noisy Biochemical Signaling. *Science* **2011**, *334*, 354–358. [CrossRef]
29. Lan, G.; Sartori, P.; Neumann, S.; Sourjik, V.; Tu, Y. The energy-speed-accuracy trade-off in sensory adaptation. *Nat. Phys.* **2012**, *8*, 422–428. [CrossRef]
30. Mehta, P.; Schwab, D.J. Energetic costs of cellular computation. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17978–17982. [CrossRef]
31. Cao, Y.; Wang, H.; Ouyang, Q.; Tu, Y. Biochemical oscillations. *Nature Physics* **2015**, 1–8. [CrossRef]
32. Milo, R.; Phillips, R. *Cell Biology by the Numbers*; Garland Science: New York, NY, USA, 2015; p. 2015.
33. Moran, U.; Phillips, R.; Milo, R. SnapShot: Key numbers in biology. *Cell* **2010**, *141*, 1262. [CrossRef]
34. Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* **2012**, *75*, 126001. [CrossRef]
35. Still, S.; Sivak, D.A.; Bell, A.J.; Crooks, G.E. Thermodynamics of Prediction. *Phys. Rev. Lett.* **2012**, *109*, 120604. [CrossRef] [PubMed]
36. Ouldridge, T.E.; Govern, C.C.; Rein, P. Thermodynamics of Computational Copying in Biochemical Systems. *Phys. Rev. X* **2017**, 021004. [CrossRef]
37. Rein, P.; Becker, N.B.; Ouldridge, T.E.; Mugler, A. Fundamental Limits to Cellular Sensing. *J. Stat. Phys.* **2016**, *162*, 1395–1424. [CrossRef]
38. Sagawa, T.; Ito, S. Maxwell's demon in biochemical signal transduction transduction with feedback loop. *Nat. Commun.* **2015**, *6*, 7498. [CrossRef]
39. Barato, A.C.; Hartich, D.; Seifert, U. Information-theoretic versus thermodynamic entropy production in autonomous sensory networks. *Phys. Rev. E* **2013**, *87*, 042104. [CrossRef]
40. Barato, A.C.; Hartich, D.; Seifert, U. Efficiency of cellular information processing. *New J. Phys.* **2014**, *16*, 103024. [CrossRef]
41. Bo, S.; Giudice, M.D.; Celani, A. Thermodynamic limits to information harvesting by sensory systems. *J. Stat. Mech. Theory Exp.* **2015**, *2015*, P01014. [CrossRef]
42. Govern, C.C.; ten Wolde, P.R. Energy Dissipation and Noise Correlations in Biochemical Sensing. *Phys. Rev. Lett.* **2014**, *113*, 258102. [CrossRef]
43. Barato, A.C.; Seifert, U. Thermodynamic Uncertainty Relation for Biomolecular Processes. *Phys. Rev. Lett.* **2015**, *114*, 158101. [CrossRef]
44. Brittain, R.A.; Jones, N.S.; Ouldridge, T.E. What we learn from the learning rate. *J. Stat. Mech. Theory Exp.* **2017**, *6*, 063502. [CrossRef]
45. Goldt, S.; Seifert, U. Stochastic Thermodynamics of Learning. *Phys. Rev. Lett.* **2017**, *118*, 010601. [CrossRef] [PubMed]
46. Parrondo, J.M.R.; Horowitz, J.M.; Sagawa, T. Thermodynamics of information. *Nat. Phys.* **2015**, *11*, 131–139. [CrossRef]
47. Becker, N.B.; Mugler, A.; ten Wolde, P.R. Prediction and Dissipation in Biochemical Sensing. *arXiv* **2013**, arXiv:1312.5625. Available online: http://arxiv.org/abs/1312.5625 (accessed on 9 December 2019).

48. Horowitz, J.M.; Esposito, M. Thermodynamics with Continuous Information Flow. *Phys. Rev. X* **2014**, *4*, 031015. [CrossRef]
49. Allahverdyan, A.E.; Janzing, D.; Mahler, G. Thermodynamic efficiency of information and heat flow. *J. Stat. Mech. Theory Exp.* **2009**, *2009*, P09011. [CrossRef]
50. Sartori, P.; Granger, L.; Lee, C.F.; Horowitz, J.M. Thermodynamic costs of information processing in sensory adaptation. *PLoS Comput. Biol.* **2014**, *10*, e1003974. [CrossRef]
51. Hartich, D.; Barato, A.C.; Seifert, U. Sensory capacity: An information theoretical measure of the performance of a sensor and sensory capacity. *Phys. Rev. E* **2016**, *93*, 022116. [CrossRef]
52. Falasco, G.; Rao, R.; Esposito, M. Information Thermodynamics of Turing Patterns. *Phys. Rev. Lett.* **2018**, *121*, 108301. [CrossRef]
53. Tkačik, G.; Walczak, A.M. Information transmission in genetic regulatory networks: A review. *J. Phys. Condens. Matter Inst. Phys. J.* **2011**, *23*, 153102. [CrossRef]
54. Tkačik, G.; Walczak, A.M.; Bialek, W. Optimizing information flow in small genetic networks. *Phys. Rev. E* **2009**, *80*, 031920. [CrossRef]
55. Walczak, A.M.; Tkačik, G.; Bialek, W. Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys. Rev. E* **2010**, *81*, 041905. [CrossRef] [PubMed]
56. Tkačik, G.; Walczak, A.M.; Bialek, W. Optimizing information flow in small genetic networks. III. A self-interacting gene. *Phys. Rev. E* **2012**, *85*, 041903. [CrossRef] [PubMed]
57. Mugler, A.; Walczak, A.; Wiggins, C. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys. Rev. E* **2009**, *80*, 041921. [CrossRef] [PubMed]
58. Rieckh, G.; Tkačik, G. Noise and Information Transmission in Promoters with Multiple Internal States. *Biophys. J.* **2014**, *106*, 1194–1204. [CrossRef] [PubMed]
59. Sokolowski, T.R.; Tkačik, G. Optimizing information flow in small genetic networks. IV. Spatial coupling. *Phys. Rev. E* **2015**, *91*, 062710. [CrossRef] [PubMed]
60. de Ronde, W.H.; Tostevin, F.; ten Wolde, P.R. Feed-forward loops and diamond motifs lead to tunable transmission of information in the frequency domain. *Phys. Rev. E* **2012**, *86*, 021913. [CrossRef] [PubMed]
61. Gregor, T.; Wieschaus, E.F.; McGregor, A.P.; Bialek, W.; Tank, D.W. Stability and nuclear dynamics of the Bicoid morphogen gradient. *Cell* **2007**, *130*, 141–152. [CrossRef]
62. Gregor, T.; Tank, D.W.; Wieschaus, E.F.; Bialek, W. Probing the limits to positional information. *Cell* **2007**, *130*, 153–164. [CrossRef]
63. Pahle, J.; Green, A.K.; Dixon, C.J.; Kummer, U. Information transfer in signaling pathways: A study using coupled simulated and experimental data. *BMC Bioinform.* **2008**, *9*, 139. [CrossRef]
64. Selimkhanov, J.; Taylor, B.; Yao, J.; Pilko, A.; Albeck, J.; Hoffmann, A.; Tsimring, L.; Wollman, R. Accurate information transmission through dynamic biochemical signaling networks. *Science* **2014**, *346*, 1370–1373. [CrossRef]
65. Mancini, F.; Wiggins, C.H.; Marsili, M.; Walczak, A.M. Time-dependent information transmission in a model regulatory circuit. *Phys. Rev. E* **2013**, *88*, 022708. [CrossRef] [PubMed]
66. Mancini, F.; Marsili, M.; Walczak, A.M. Trade-offs in delayed information transmission in biochemical networks. *J. Stat. Phys.* **2015**, *1504*, 03637. [CrossRef]
67. Kepler, T.B.; Elston, T.C. Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. *Biophys. J.* **2001**, *81*, 3116–3136. [CrossRef]
68. Raj, A.; Peskin, C.S.; Tranchina, D.; Vargas, D.Y.; Tyagi, S. Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biol.* **2006**, *4*, e309. [CrossRef] [PubMed]
69. Friedman, N.; Cai, L.; Xie, X. Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression. *Phys. Rev. Lett.* **2006**, *97*, 168302. [CrossRef] [PubMed]
70. Walczak, A.M.; Sasai, M.; Wolynes, P.G. Self-consistent proteomic field theory of stochastic gene switches. *Biophys. J.* **2005**, *88*, 828–850. [CrossRef] [PubMed]
71. Cai, L.; Friedman, N.; Xie, X.S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **2006**, *440*, 358–362. [CrossRef]
72. So, L.h.; Ghosh, A.; Zong, C.; Sepúlveda, L.A.; Segev, R.; Golding, I. General properties of the transcriptional time-series in E. Coli. *Nat. Genet.* **2011**, *43*, 554–560. [CrossRef]

73. Desponds, J.; Tran, H.; Ferraro, T.; Lucas, T.; Dostatni, N.; Walczak, A.M. Precision of Readout at the hunchback Gene: Analyzing Short Transcription Time Traces in Living Fly Embryos. *PLoS Comput. Biol.* **2016**, *12*, e1005256. [CrossRef]
74. Cover, T.; Thomas, J. *Elements of Information Theory*; John Wiley: New York, NY, USA, 1991.
75. Levine, M.V.; Weinstein, H. AIM for Allostery: Using the Ising Model to Understand Information Processing and Transmission in Allosteric Biomolecular Systems. *Entropy* **2015**, *17*, 2895–2918. [CrossRef]
76. Cuendet, M.A.; Weinstein, H.; Levine, M.V. The Allostery Landscape: Quantifying Thermodynamic Couplings in Biomolecular Systems. *J. Chem. Theory Comput.* **2016**. [CrossRef] [PubMed]
77. Crooks, G.E. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *J. Stat. Phys.* **1998**, *90*, 1481–1487. [CrossRef]
78. Tome, T.; de Oliveira, M.J. Entropy Production in Nonequilibrium Systems at Stationary States. *Phys. Rev. Lett.* **2012**, *108*, 020601. [CrossRef]
79. Hornos, J.E.M.; Schultz, D.; Innocentini, G.C.P.; Wang, J.; Walczak, A.M.W.; Onuchic, J.N.; Wolynes, P.G. Self-regulating gene: An exact solution. *Phys. Rev. E* **2005**, 1–5. [CrossRef]
80. Miekisz, J.; Szymanska, P. Gene Expression in Self-repressing System with Multiple Gene Copies. *Bull. Math. Biol.* **2013**, 317–330. [CrossRef]
81. Crisanti, A.; Puglisi, A.; Villamaina, D. Nonequilibrium and information: The role of cross correlations. *Phys. Rev. E* **2012**, 061127. [CrossRef]
82. Puglisi, A.; Pigolotti, S.; Rondoni, L.; Vulpiani, A. Entropy production and coarse graining in Markov processes. *J. Stat. Mech. Theory Exp.* **2010**, 05015. [CrossRef]
83. Busiello, D.M.; Hidalgo, J.; Maritan, A. Entropy production for coarse-grained dynamics. *arXiv* **2019**, arXiv:1810.01833v2.
84. Xiong, W.; Ferrell, J.E., Jr. A positive feedback based bistable memory module that governs a cell fate decision. *Nature* **2003**, *426*, 460–465. [CrossRef]
85. Tanaka, K.; Augustine, G.J. A Positive Feedback Signal Transduction Loop Determines Timing of Cerebellar Long-Term Depression. *Neuron* **2008**, *59*, 608–620. [CrossRef]
86. Guisbert, E.; Herman, C.; Lu, C.Z.; Gross, C.A. A chaperone network controls the heat shock response in *E. coli*. *Genes Dev.* **2004**, 2812–2821. [CrossRef] [PubMed]
87. Lahav, G.; Rosenfeld, N.; Sigal, A.; Geva-zatorsky, N.; Levine, A.J.; Elowitz, M.B.; Alon, U. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.* **2004**, *36*, 147–150. [CrossRef] [PubMed]
88. Tyson, J.J.; Novák, B. Models in biology: Lessons from modeling regulation of the eukaryotic cell cycle. *BMC Biol.* **2015**, 1–10. [CrossRef] [PubMed]
89. Lucas, T.; Tran, H.; Perez Romero, C.A.; Guillou, A.; Fradin, C.; Coppey, M.; Walczak, A.M.; Dostatni, N. 3 minutes to precisely measure morphogen concentration. *PLoS Genet.* **2018**, *14*, e1007676. [CrossRef]
90. Sagawa, T.; Ueda, M. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.* **2012**, *109*, 1–5. [CrossRef]
91. Sagawa, T.; Ueda, M. Nonequilibrium thermodynamics of feedback control. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2012**, *85*, 1–16. [CrossRef]
92. Raser, J.; O'Shea, E. Control of stochasticity in eukaryotic gene expression. *Science* **2004**, *304*, 1811–1814. [CrossRef]
93. Walczak, A.M.; Onuchic, J.N.; Wolynes, P.G. Absolute rate theories of epigenetic stability. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18926–18931. [CrossRef]
94. Puglisi, A.; Villamaina, D. Irreversible effects of memory. *EPL (Europhys. Lett.)* **2009**, *88*, 30004. [CrossRef]