Scuola Internazionale Superiore di Studi Avanzati

Ph.D. Thesis

Information theory in biochemical regulatory networks: a theoretical study

Author:
Francesca Mancini

Supervisors:
Prof. Matteo Marsili
Dr. Aleksandra Walczak

May 2015

"E quindi uscimmo a riveder le stelle." (Dante Alighieri)

SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

Abstract

Physics and Chemistry of Biological Systems

Doctor of Philosophy

Information theory in biochemical regulatory networks: a theoretical study

by Francesca Mancini

In this Thesis we consider the optimization of information transmission as a viable design principle for biochemical networks. We apply this principle to a simple model regulatory circuit, given by an input and a delayed output that switch randomly between two states in continuous time.

First we maximize the transmitted information in the network at a given output delay, when the system has no external constraints and it is in steady state or can optimize its initial condition. We find that optimal network topologies correspond to common biological circuits linked to stress response and that circuits functioning out of steady state may exploit absorbing states to be more informative than in steady state.

We then take into account that biological regulatory networks need to dissipate energy in order to transmit biochemical signals and that such signaling often happens in challenging environmental conditions. Hence we explore the system's trade-offs between information transmission and energetic efficiency. At fixed delay and dissipated energy, we determine the most informative networks both in the absence and in the presence of feedback. We find that negative feedback loops are optimal at high dissipation, whereas positive feedback loops become more informative close to equilibrium conditions. Moreover, feedback allows the system to transmit almost the maximum available information at a given delay, even in the absence of dissipation.

Finally, within a game-theoretic maximin approach, we ask how a biochemical network should be constructed to be most informative in the worst possible initial condition set by the environment. We find that, in the limit of large energy dissipation, the system tunes the ratio of the input and output timescales so that the environmental disturbance is marginalized as much as possible.

Acknowledgements

This is the part I have always liked most in a Thesis, because it gives the reader a hint of the "human side" of the work. In my case this part is also undoubtedly necessary, since all of this would not have been possible without the patient and tenacious contribution of some people.

I need to thank the Committee of the Molecular and Statistical Biophysics group, for granting me the possibility to pursue the PhD while working in a company in another city. Of course I wish to thank my internal supervisor Matteo Marsili, whose aplomb has proven to be extremely valuable in some occasions during these four years. My deepest gratitude goes to my external supervisor Aleksandra Walczak: she has been like a tough elder sister, compassionate and demanding at the same time. Her commitment to our work has been an inspiration and a guidance for me.

I also want to thank the people who shared my life in Trieste, before, and in Milan, afterwards. In particular my thoughts go to my colleagues at the office, who have encouraged me to strive for this achievement in the last months. Life at work would have been much more difficult without their understanding.

Both my "natural" and "acquired" families have provided me with an unconditional moral support: mum, dad, brother, grandma, aunt and uncle, Natascia, Gianni, Dimitri, Laura, thank you all. Luckily and hopefully this is the last Thesis of my life!

Finally, as usual, I thank the partner of my life. No one can understand and share my world as he does. Our story has started together with our PhD and has endured its ups and downs, making us stronger.

I am leaving the research career with no regrets, I will always treasure these years in academia for giving me the tools, or at least the patience, for tackling any professional and interpersonal challenge. Whatever life brings me, I wish I will always keep the wondering and youthful attitude of a scientist.

Contents

Introduction

"Statistically, the probability of any one of us being here is so small that you'd think the mere fact of existing would keep us all in a contented dazzlement of surprise. We are alive against the stupendous odds of genetics, infinitely outnumbered by all the alternates who might, except for luck, be in our places. Even more astounding is our statistical improbability in physical terms. The normal, predictable state of matter throughout the universe is randomness, a relaxed sort of equilibrium, with atoms and their particles scattered around in an amorphous muddle. We, in brilliant contrast, are completely organized structures, squirming with information at every covalent bond." (Lewis Thomas, The Lives of a Cell)

Motivation

Today the common thought is that most of the open questions in biology should be answered by experiments. In physics the approach is different, since the theoretical and experimental approaches are considered as partners which equally contribute to scientific progress. The general theoretical principles in physics, besides explaining what has been already observed, provide a starting point to investigate what "has" to be seen. Therefore, inspired by the beautiful book of Bialek [1], we could ask ourselves

- can we reach in biophysics the same level of predictive power that is ubiquitous in other branches of physics?
- can we harmonize the physicists' quest for universal theoretical principles with the incredible diversity of living organisms?
- once such theories are identified, could they be proved meaningful in many different experimental circumstances and at the same time be expressed by simple and general principles?

Indeed, many diverse biological phenomena, such as the development of an embryo or the adaptation of bacteria to different environments, trigger physical fundamental questions. But how do we frame them into theories?

A starting point could be to observe that life is a macroscopic state which is generated and maintained by microscopic mechanisms such as DNA replication and protein synthesis. While we have made a lot of progress in studying the microscopic level, we still have to identify the "order parameters" that identify the macroscopic level. The identification of such order parameters is difficult, however it can be tackled by making quantitative models of specific biological systems and grouping them into different regions of the parameter space. Real and functional biological systems would realistically belong only to a restricted region of the parameter space. If there is a *principle* that is able to single out such region, then we can use it to build a theory and to calculate a priori the properties of biological systems.

A viable principle is to postulate that biological systems find *optimal solutions* for the tasks they need to accomplish, or in other words they *maximize* specific objective functions. This approach is familiar to physicists, indeed many theories are formulated as variational principles, such as least action in classical and quantum mechanics or the minimization of free energy in thermodynamics. Nevertheless, it could be objected that living systems are still in the evolutionary flow, constantly climbing the "fitness gradient", and are thus far from having reached an optimal performance. So our attempts to theorize biological systems by maximizing a function might be wrong, not because we chose the wrong function, but more profoundly because life is not at the extreme of *any function*.

Still we can support our claim of the existence of an *optimization principle* by making it meaningful and predictive. First of all we should be able to measure our chosen objective function F and compare it to its theoretical maximum, constrained by physical limits. Then maximizing F should allow us to make predictions testable in different experiments. We should restrict ourselves to systems where we have direct evidence for the maximization of F. But even when the systems are not at the maximum of F, our approach could still be an useful idealization.

Recently many candidates for the objective function F have been taken into consideration [2–8]. Possible choices are minimization of biochemical noise [4, 8, 9], minimization of losses in case of unknown external signals [6], maximization of positional information [10] or optimization of resources [11]. Some of these approaches have taken into account evolutionary effects [3, 10].

It has also been argued that the important process in biology is the *flow of information*. Living organisms collect information from the environment or from their internal state so as to make choices. In this Thesis we choose to explore the principle that *biological systems maximize information transmission*, or the efficiency with which information is represented [12]. From the mathematical point of view, Shannon [13] proved that there

is only one precise measure of information, which is connected with statistical mechanics. From the experimental point of view, such principle has been tested in the embryonic development of the fruitfly [5].

Here we do not test the principle but rather assume it as a matter of fact. We apply it to a simple biochemical network in order to devise its maximally informative topologies. We first allow our system to behave "ideally" and then ask how adding energetic constraints or imposing unfavorable initial conditions changes the system's optimal response.

Outline of the thesis

Here follows an outline of the Thesis: in Chapter 1 we first provide a biological motivation for the choice of optimal information transmission as a design principle. We then briefly revise the concept of Shannon mutual information as a measure of correlation between two variables, together with its definition and its main properties. After a short review of the literature on the topic, we introduce our simple model of a biochemical regulatory network: it consists of an input and a delayed output, which switch randomly in continuous time between an "active" and an "inactive" state. Such states could describe the presence/absence of a sugar source in the environment, as an input to the network, and the presence/absence of its degrading enzyme, as the network output, respectively. In sections 1.4.2.2, 1.4.2.3 and 1.4.2.4 we look for optimally informative network topologies when the system is in steady state, both in the absence and in the presence of feedback. Then in section 1.4.2.5 we ask how the optimal solutions change when the system is allowed to optimize its initial condition. In all the above cases we look for maximal information transmission at a fixed delay of the output response.
In Chapter 2 we take into account energetic constraints. While the model regulatory network in Chapter1 was allowed to function extremely far from equilibrium, and thus to dissipate an infinite amount of energy, it is now constrained to optimize the transmitted information by working close to equilibrium. We start with a quick summary of stochastic thermodynamics and, after mentioning relevant references, we endow the "ideal" model of Chapter 1 with an energetic framework. In sections 2.3.2 and 2.3.3 we analyze the system in steady state in the absence and presence of feedback. Then in section 2.3.4 we use the obtained results to estimate the energetic efficiency of information transmission in the two-component osmoregulatory network EnvZ-OmpR of *E. coli*.
Finally, in Chapter 3 we consider the realistic event where the system is forced to maximize the transmitted information in the worst possible initial conditions. We use a game-theoretic approach to model the interaction between the system and the environment as a maximin game, where the first player wants to maximize the transmitted

information at all times, while the second player wants to minimize it.
We end the Thesis with some concluding remarks and future perspectives.

The material presented in this Thesis is mainly based on the following publications:

- F. Mancini, C. H. Wiggins, M. Marsili and A. M. Walczak, *Time-dependent information transmission in a model regulatory circuit*, Physical Review E 88, 022708 (2013)
- F. Mancini, M. Marsili and A. M. Walczak, *Trade-offs in delayed information transmission in biochemical networks*, arXiv:1504.03637 [q-bio.MN] (2015)

Chapter 1

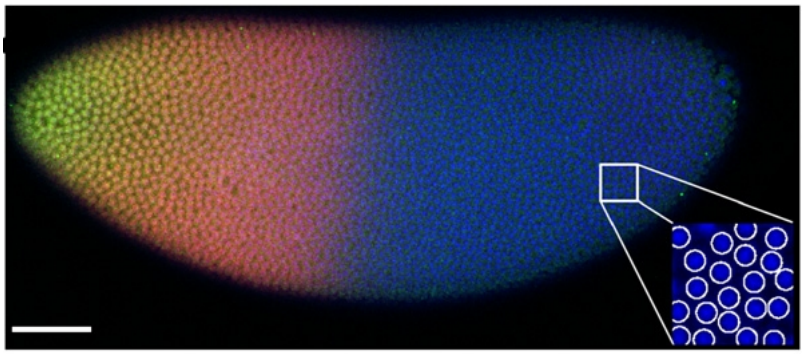
Optimizing information transmission

1.1 Does Biology care about information?

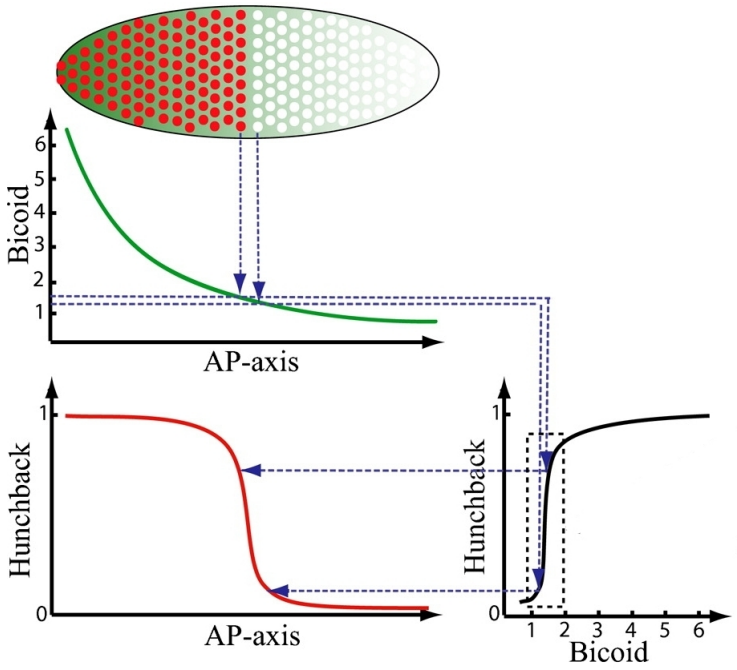
The optimization principle of *maximal information transmission* has been successfully applied to the early embryonic development of *D. Melanogaster* [14]. This organism is a case study for *spatial patterning*: nuclei in the embryo, even if equipped with the same DNA, differentiate into distinct parts of the body by modulating their gene expression. The precise and reproducible spatial domains of differential gene expression are possible thanks to few maternal chemical cues, called *mathernal morphogens* [15]. In particular, a maternal mRNA coding for the morphogen Bicoid is laid in the anterior part of the embryo. After translation, the morphogen diffuses towards the back of the embryo, establishing a smooth gradient along the anterior-posterior (AP) axis (see Fig.1.1). Each nucleus reads the local concentration of Bicoid (and other morphogens) and subsequently modulates the expression of four downstream genes, called *gap genes*. One of the gap genes is *hunchback*: as shown in Fig.1.1b, its expression has a sharp sigmoidal response to Bicoid concentration, similar to a binary switch [16, 17].

Thanks to simultaneous measurements [17] of the concentrations of Bicoid (c) and Hunchback (g) (see Fig.1.1a), Tkačik and collaborators [5] were able to sample the joint probability distribution $P(c, g)$ and compute from that the experimental mutual information $I_{\text{exp}}(c; g) = 1.5 \pm 0.1$ bits [1]. This value is higher than what one would expect from a truly binary switch, which conveys 1 bit of information. It is therefore instructive to have another measure to compare 1.5 bits to.

[1] The mathematical definition of mutual information will be provided in the subsequent section, here we just want to provide an example of its biological applicability.



(a)



(b)

Figure 1.1: taken from [17] and modified. (b) In the early development of *D. melanogaster* embryo, a smooth gradient of Bicoid concentration along the anterior-posterior (AP) axis is translated into a sharp boundary of Hunchback expression. (a) Scanning confocal microscope image of a *D. melanogaster* embryo in early nuclear cycle 14, stained for DNA (blue), Hunchback (red), and Bicoid (green); scale bar $50\mu m$. Inset shows how DNA staining allows for automatic detection of nuclei.

To do so, Tkačik et al. computed the theoretical maximum for the information transmitted between Bicoid and Hunckback (also known as *channel capacity*). They started from rewriting the joint probability $P(c, g)$ as $P(c, g) = P(g|c)P(c)$. The conditional probability $P(g|c)$ of Hunchback expression given the Bicoid signal can be computed experimentally and approximated with a Gaussian $\mathcal{G}(g; \bar{g}(c), \sigma_g(c))$, where $\bar{g}(c)$ is the mean response of the regulatory element and $\sigma_g^2(c)$ the noise in the response. Holding $P(g|c)$ fixed and optimizing $P(c)$ numerically, yielded the channel capacity of $I^*(c; g) = 1.7$

bits. Such finding shows that the real biological system is operating close to what is achievable given the noise, that is $I_{\text{exp}}(c;g)/I^*(c;g) \approx 90\%$. This provides a good motivation for taking maximization of information transmission as a viable design principle. Additionally, the optimal probability distribution of Bicoid $P^*(c)$ predicts the optimal distribution of Hunchback expression levels through $P^*(g) = \int dc P(g|c)P^*(c)$. As it is shown in Fig.1.2, the predicted probability $P^*(g)$ (red) matches the experimental one $P_{\text{exp}}(g)$ (black) extremely well.

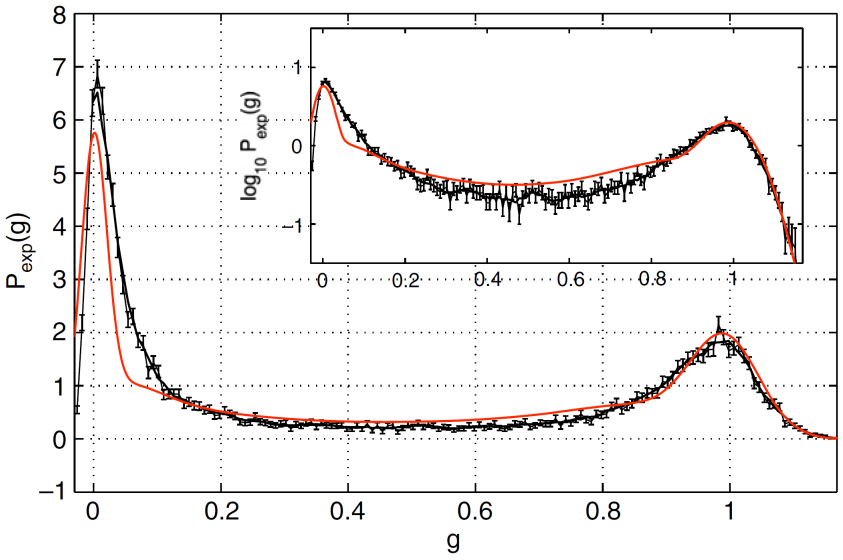


Figure 1.2: taken from [5]. The measured (black) and predicted optimal (red) distribution $P(g)$ of Hunchback expression levels across an ensemble of nuclei in *D. melanogaster* embryo. The expression level g goes from 0 (no induction, posterior) to 1 (full induction, anterior). A significant fraction ($\approx 30\%$) of the nuclei express an intermediate level of Hunchback which, given the low noise in the system, could represent a third distinct expression level, besides 0 and 1. This would agree with the observed information of 1.5 bits, that intuitively corresponds to $2^{1.5} \approx 3$ distinguishable expression levels.

1.2 Entropy and Mutual information

Having built some intuition about the application of mutual information to biological systems, we can now provide its formal definition and its main properties (the interested reader can refer to [18, 19] for a detailed review of the subject). We will indicate the input signal with z and the output signal with x, and we will consider both signals to be *discrete* (however, all the following considerations are valid also for continuous variables).

Since biochemical networks are intrinsically noisy, the mapping of the input z to the output x is a probability distribution $P(x|z)$. If the input itself is drawn from a distribution $P(z)$, then pairs of input/output measurements are jointly distributed according to

$$P(z,x) = P(x|z)P(z). \tag{1.1}$$

We are interested in measuring how strongly input z and output x are dependent on each other. In other words, we want to find a measure that would quantify "how much" one can know, in principle, about the value of z by receiving its output x, given a certain input/output relation $P(x|z)$ and some input distribution $P(z)$.
The first quantity that one thinks of is the covariance

$$Cov(z,x) = \sum_z \sum_x (z-\bar{z})(x-\bar{x})P(z,x),$$

where $\bar{z}, \bar{x}$ indicate the averages of z and x, respectively. However, it is easy to think of situations where z and x are statistically related but the covariance is 0. Indeed, covariance (or correlation coefficient) only measures the *linear* dependence between variables. We would like instead our measure of interdependency to be very general, free of assumptions on the probability distribution underlying the data. According to Shannon [13], there is a *unique assumption-free* measure of interdependency, called *mutual information* $I[z,x]$ between z and x (see Fig.1.3).

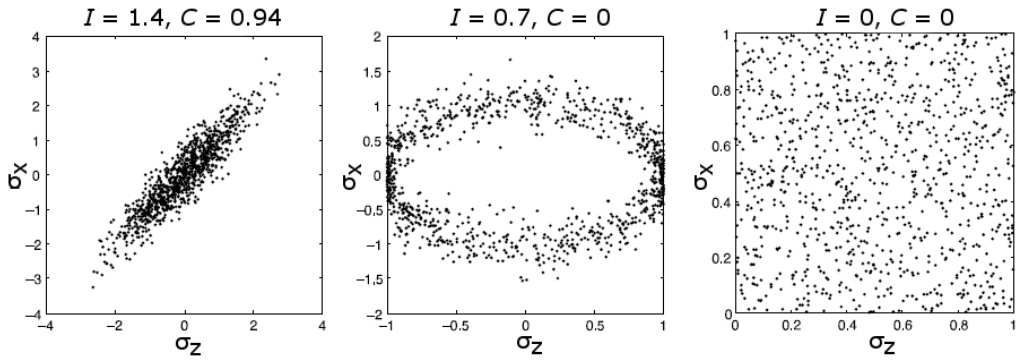


Figure 1.3: (taken from [12] and modified). Examples of two variables, drawn from three joint distributions. On the left, the variables are linearly correlated, and both the correlation coefficient and the mutual information are close to 1. In the middle, the variables are interdependent, but not in a linear sense. The correlation coefficient is 0, but mutual information gives a non-zero value. On the right, the variables are statistically independent, and both linear correlation and mutual information give zero signal.

Let us first recall the definition of *Shannon entropy*

$$S[P(x)] = -\sum_x P(x)\log_2 P(x), \tag{1.2}$$

as a measure of *uncertainty* of the output x, whose distribution is $P(x)$. Given the input/output relation $P(x|z)$ we can also use the Shannon entropy

$$S[P(x|z)] = -\sum_x P(x|z)\log_2 P(x|z) \tag{1.3}$$

to indicate the uncertainty in x if we know its input z. If we compute the difference of the two entropies 1.2 and 1.3, we can measure how much our uncertainty about x *has been reduced* by knowing z: $\Delta S = S[P(x)] - S[P(x|z)]$. We can repeat the measure of ΔS for different concentrations of input z and take an average according to its distribution $P(z)$. The resulting quantity is our mutual information:

$$I[z,x] = \sum_z P(z)(S[P(x)] - S[P(x|z)]). \tag{1.4}$$

This quantity is a scalar number expressed in *bits* and measures how much, *on average*, our uncertainty in a variable (x in this case) is decreased by knowing the value of a related variable (e.g. z). Using equations 1.1,1.2 and 1.3, we can reformulate the mutual information as

$$\begin{aligned} I[z,x] &= \sum_z\sum_x P(z,x)\log_2\frac{P(z,x)}{P(z)P(x)} &(1.5)\\ &= \sum_z P(z)\sum_x P(x|z)\log_2\frac{P(x|z)}{P(x)} &(1.6)\\ &= \sum_x P(x)\sum_z P(z|x)\log_2\frac{P(z|x)}{P(z)}. &(1.7)\end{aligned}$$

From the above expressions the following properties follow:

- information is a symmetric quantity in z and x: the information the input has about the output is the same as the information the output has about the input.
- if the joint distribution of inputs and outputs factorizes, $P(z,x) = P(z)P(x)$, then $I(z;x) = 0$. In this case, the entropy of the whole system would be the sum of the individual entropies: $S[P(z,x)] = S[P(z)] + S[P(x)]$. On the other hand, if the variables are not independent, the entropy of the system is reduced by the mutual information, $S[P(z,x)] = S[P(z)] + S[P(x)] - I[z,x]$, and mutual information is bounded from above by the individual entropies, $I[z,x] < \min(S[P(z)], S[P(x)])$.

Moreover, mutual information

- can be defined both for continuous and discrete variables,

- is reparametrization invariant: $I[z, x] = I[f(z), h(x)]$, where $f(z)$ and $h(x)$ are one-to-one functions of their arguments,
- obeys the data processing inequality: given a noisy mapping $z \to x \to k$, then $I[z, k] \leq I[z, x]$, that is information either decreases or stays the same, but it is never "created" in the noisy transmission process,
- has a clear interpretation: if there are $I[z, x]$ bits of mutual information between z and x, there are $2^{I[z,x]}$ distinguishable levels of x obtainable by giving z.

1.3 Making models

To respond to environmental changes, regulatory biochemical networks need to transform the molecular signals they receive as input into concentrations of response molecules. These processes are inherently stochastic, as both input and output molecules are often present in small numbers. This observation has motivated a number of recent works which pose the search for design principles as an optimization problem over the network topology and reaction rates. Not all designs of biochemical networks are equally represented in cells [20], which raises the question of whether the prevalence of particular network motifs arises because they are best suited for specific tasks the cell has to fulfill. In order to attempt to answer this question, and to explore the functions and limitations of given network architectures, one can consider a well defined objective function, such as rapidity of response [20, 21], minimization of noise [8, 22] or information transmission between the input and output [12, 23–25], and compare the performance of particular circuits under a set of constraints (e.g. molecular cost, noise). Phrasing the problem as an optimization over the parameters and probability distributions of these networks allows one to find the optimal circuit that best fits this one specific function. The optimal architecture corresponds to how one would "design" or "build" a circuit if the objective was to satisfy a known function.

As previously discussed, a possible objective function could be the optimal information transmission between a biochemical input and its output. The idea of information optimization in biomolecular networks was tested in early development of *D. melanogaster* embryo (see section 1.1). More recently, the information transmission measured in an inflammatory cytokine signaling pathway of murine fibroblasts [26] was used to find a tree-like network topology, thus allowing to identify the bottlenecks that limit signal integration. This combined theoretical and experimental approach singled out the conditions under which feedback, time integration and collective cell response can increase information transmission. The NF-kappa B and ERK pathways were recently used to

demonstrate that dynamical measurements of the response can transmit more information than static or one time readouts [27]. Lastly, an information-theoretic approach was used in an experimental and numerical study to show the interdependence of stochastic processes controlling enzymatic calcium signaling under different cellular conditions [28].

Theoretical studies based on the optimization of information transmission in regulatory circuits have demonstrated that, within a network that functions at steady state, the system can exploit the molecular details of the network to transmit information while paying a molecular cost [23, 29]. Positive feedback was shown to increase the effective integration time and average out input fluctuations, thus allowing for reliable information transmission at low input concentrations [25]. On the contrary, negative feedback reduces noise at high input concentrations by reducing the effective nonlinearity of the input-output relation. Molecular strategies, such as using feedforward loops [24] and slow binding/unbinding dynamics [30], also increase information transmission, because they lead to a quasi-binary readout and multimodal output distributions.

In many situations biochemical signals change with time, which has led to an interest in the information-optimal response to pulses in signaling cycles [31] and to oscillatory driving [32]. Similarly to what was found for stationary signals, circuits that produce bimodal output states [32] transmit more information. Tostevin and ten Wolde looked at time dependent Gaussian processes in a linearized regulatory circuit and found that those network properties that are important for transmitting information about instantaneous signals may not be those that are relevant for information transmission of entire trajectories [33, 34]. Within the same framework de Ronde and colleagues focused on understanding the role of regulation and found that positive feedback increases the fidelity of low frequency signals, whereas negative feedback proves better at high frequencies [35]. In feedforward circuits, they showed that topologies alone are not sufficient to characterize network properties, but that interaction strength also plays an important role [36].

One should also take into account that regulatory response is often at a delay relative to input signaling, because of e.g. transcription and translation processes, cellular compartmentalization, etc [37]. Examples include the chemotactic response of bacteria [38] or amoeba [39] to nutrients or conversely to antibiotics [40]. This delay is intrinsic and unavoidable in a biological circuit - mRNAs and proteins are not produced instantaneously and interpreting the initial signal takes time. Optimal design therefore entails maximizing information transmission between the input at a given time and the output at a later time. Nemenman [41] has shown that, in simple regulatory circuits, there is an optimal time lag for which the mutual information between the input and the delayed output is maximized.

1.4 A model regulatory circuit for time-dependent information transmission

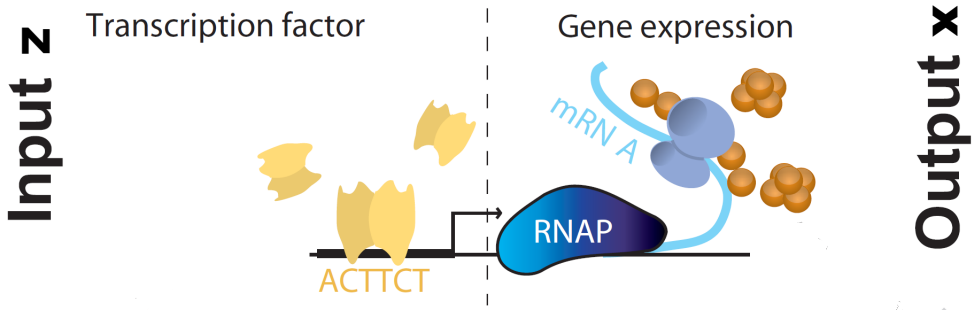


Figure 1.4: Cartoon of a simple biochemical regulatory network with an input z (e.g. transcription factor) and an output x (e.g. gene expression). Original picture taken from [5] and modified.

Given a fixed unavoidable delay in the response, we ask what is the optimal way to design a regulatory system to optimally transmit information between an input at an initial time and an output at a later time. In our approach the delay is an intrinsic property of the system, which we cannot control, but rather treat as an additional constraint. Given this constraint, we ask about the optimal parameters and architecture of the system that maximally transmits information between the input and the output. To find this optimal system, we optimize over both the initial distribution and the elements of the network.

Our goal is to focus on how the natural cellular delays constrain the architectures of regulatory circuits. We find that the design of circuits that optimally transmit information between the input and a delayed output corresponds to known circuit's architectures (push-pull networks) [42–44]. Moreover, we find that this prediction is robust, because optimal architectures do not change with the length of the imposed delay.

Our strategy is to address these issues within the simplest possible model, composed of two binary components, z and x, that switch randomly in continuous time. This can model, e.g., the binding/unbinding of a transcription factor z to its corresponding gene x, which is subsequently activated/inactivated (see Fig. 1.4). The model sacrifices molecular details of biochemical regulatory networks, which can be very complex and whose features can have an effect on their information processing functions. Yet, the simplicity of our model makes our approach and the resulting results as clear as possible.

1.4.1 Model description

We consider a system of two dynamical variables x, z that describe two genes, or a gene and a protein. Each one of them switches between an activated $(+1)$ and inactivated (-1) state according to the rates defined in Fig. 1.5.

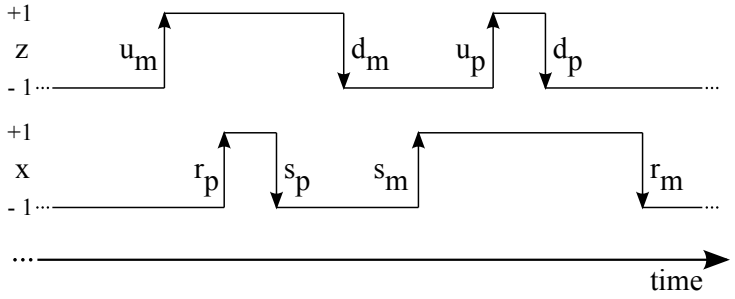


Figure 1.5: Time evolution of random variable $z(t)$, which models a biochemical input transitioning from/to a down-regulated state (-1) to/from an up-regulated state $(+1)$, with rates $\{u_m, u_p\}/\{d_m, d_p\}$, respectively. Random variable $x(t)$ models activation $(+1)$ or deactivation(-1) of a biochemical output: it is regulated by z, with which it aligns ('activation', or up-regulation) with rates r_m or r_p or anti-aligns ('repression', or down-regulation) with rates s_m or s_p. The subscripts m and p in the rates account for the state of the other variable, that is -1 and $+1$, respectively.

Specifically, x aligns or anti-aligns with z with rates r or s, respectively. The subscripts m and p in the rates indicate the state of the other variable (-1 or $+1$) at the moment of the flip of the first variable. Our system at any time is fully described by a four-state probability distribution $P(y)$, where the state y is a joint variable for the output and the input and it is defined as

$$y{=}(x, z) \in \{(-,-),(-,+),(+,-),(+,+)\}. \tag{1.8}$$

The temporal evolution of the conditional probability $P(y'|y)$ to find the system in state y' at time t given state y at $t{=}0$ is given by the continuous-time master equation

$$\frac{dP}{dt} = -\mathcal{L}P, \tag{1.9}$$

where $\mathcal{L}$ is a 4×4 transition matrix set by the rates of switching between the states (shown in Fig. 1.5 and constrained to be in the range $[0, 1]$):

$$\mathcal{L} = \begin{pmatrix} u_m + s_m & -d_m & -r_m & 0 \\ -u_m & d_m + r_p & 0 & -s_p \\ -s_m & 0 & u_p + r_m & -d_p \\ 0 & -r_p & -u_p & d_p + s_p \end{pmatrix}. \tag{1.10}$$

In $P(y'|y)$ and in the rest of the section, primed variables refer to the system state at time t, and unprimed variables to the initial time 0. The solution of the master equation can be formally written as $P{=}e^{-t\mathcal{L}}$ and is conveniently expressed in terms of its expansion in left and right eigenvectors of $\mathcal{L}$ (see Appendix A.1). In particular, the (normalized) right eigenvector corresponding the null eigenvalue is the **stationary state** $P_\infty(y'){=}\lim_{t\to\infty} P(y'|y)$.

As explained above, for the sake of realism we want to take into account the intrinsic delays with which biochemical regulatory networks respond to signals. Therefore we compute the mutual information between the input z at time 0 and the output x' at a time delay t, which is defined as (see section 1.2)

$$I[x_t, z_0] = \sum_{x',z} P(x', z) \log_2 \frac{P(x', z)}{P(x')P_0(z)}. \tag{1.11}$$

The joint probability distribution $P(x', z)$ can be readily derived from the conditional distribution $P(y'|y)$ and the initial distribution $P_0(y)$ (see Appendix A.1).

Intuitively, a system that conveys most information between the input and output requires matching the properties of the network (in the case of our model defined by the switching rates) with the properties of the input and output distributions. The maximum information transmitted by a system, termed the capacity of this system [13, 18], is defined as the optimum of Eq. 1.11 with respect to the input distribution. We discuss two cases: in the first one the system is in steady state, e.g. the response to a morphogen gradient such as FGF in development [45]. In the second case the initial state of the system is not the steady state, e.g. production of an enzyme to metabolize a newly available sugar [37]. While in the former the initial state of the regulatory network is determined by $\mathcal{L}$ (i.e. $P_0(y) = P_\infty(y)$), in the latter $P_0(y)$ may also be optimized. We consider both optimization cases in the context of our model. A trivial way to maximize information transmission is for the input to change infinitely slowly relative to a fixed delay time t (i.e. $u_l, d_l \to 0$, with $l = p, m$ following the notation of Fig. 1.5), such that for any finite t, the output yields a noiseless readout of the input, i.e. $P(x_t = z_0) \approx 1$. In short, nothing happens. Instead, we are interested in regulatory responses to changes in the input.

To constrain our optimization such that information is transduced on a timescale set by the system's own dynamics, we optimize the quantity

$$\mathcal{I}(\tau) = I[x_{t=\tau/\lambda}, z_0], \tag{1.12}$$

where the rate λ is given by the smallest non-zero eigenvalue of $\mathcal{L}$ and is the inverse of the system's largest relaxation timescale.
Such time rescaling allows us to keep our model as general as possible and to avoid making questionable choices for the units of measure. Measuring time in seconds, in minutes or in hours would restrict our model's applicability only to the biochemical networks which operate on that timescales.
For the sake of clarity, we stress that the delay τ is an *extrinsic* property of the system: it is not related to one of the network rates but, at the same time, it is not due to downstream readout. An interpretation of it could be the time needed for the sythesis of an output protein x after the binding to DNA of a transcription factor z. The intermediate step of mRNA synthesis is time consuming and unavoidable, so that when the input changes, the output simply cannot – is extrinsecally forbidden to – change instantaneously or before τ.
However, since delay τ is measured in terms of the relaxation time of the system $1/\lambda$, it is comparable to the system's own dynamics. So the overall delay in real time t depends on the values of the rates.

We remove arbitrariness in the choice of time units by fixing the magnitude of the maximum rate to be 1. $\mathcal{I}(\tau)$ implicitly depends on the rates appearing in $\mathcal{L}$ and, if it is not set by $P_0(y) = P_\infty(y)$, on the initial distribution $P_0(y)$. We find network architectures that maximize $\mathcal{I}$ in Eq. (1.12) for each rescaled delay $\tau = \lambda t$ over the rates (code that performs the optimization is available at http://infodyn.sourceforge.net).

To summarize our procedure, for a fixed intrinsic delay of the output, measured in units of the relaxation time of the system, we look for the optimal rates of the network defined in Fig. 1.5. We then scan the delay time to see how the properties of the optimal networks change (see Appendix A.2 for details).

1.4.2 Results

1.4.2.1 A general expression for mutual information

Since our model describes a 1-bit symmetric system, we can write the mutual information in a general and intuitive form, related to the entropy of an effective two-state spin variable.
Consider any arbitrary distribution $P(a, b)$ where $\{a, b\} \in \{-1, +1\}$ and the distribution enjoys symmetry under flipping $-1 \leftrightarrow +1$.
In this case $P(+,+) = P(-,-)$ and $P(+,-) = P(-,+)$.

Let us define $-1 < \mu < 1$ such that

$$P(+,+) = P(-,-) = \frac{1+\mu}{4}, \tag{1.13}$$

$$P(-,+) = P(+,-) = \frac{1-\mu}{4}. \tag{1.14}$$

From these, we see that $P(a) = P(b) = 1/2$ and

$$\begin{aligned} I[a,b] &= \sum_{a,b} P(a,b) \log_2 \frac{P(a,b)}{P(a)P(b)} \\ &= \sum_{a,b} P(a,b) \log_2 4P(a,b) \\ &= 2P(-,-) \log_2 4P(-,-) + 2P(-,+) \log_2 4P(-,+) \\ &= \frac{1}{2}(1+\mu)\log_2(1+\mu) + \frac{1}{2}(1-\mu)\log_2(1-\mu), \end{aligned} \tag{1.15}$$

where μ can be interpreted as the "effective magnetization" of the spin variable [2].

1.4.2.2 Simple activation

In order to gain intuition, we start by considering the simplest possible regulatory network, which will be referred to as **model** A : here z up-regulates x perfectly, symmetrically, and without feedback. In other words, x is slaved to z and switches only if $x \neq z$ but, due to the stochasticity of the model, it may not align immediately with z.
Hence the rates of Fig. 1.5 are set in the following way

- $u_m = u_p = d_m = d_p \equiv u$,
- $r_m = r_p \equiv r = 1$,
- $s_m = s_p = 0$,

which leaves us with only two timescales, related to u and r. The initial probability $P_0(y)$ is set equal to the steady state $P_\infty(y)$.
In this case the transition rate matrix $\mathcal{L}$ is given by

$$\mathcal{L} = \begin{pmatrix} u & -u & -r & 0 \\ -u & u+r & 0 & 0 \\ 0 & 0 & u+r & -u \\ 0 & -r & -u & u \end{pmatrix} \tag{1.16}$$

[2] Note also that $P(a=b,b) - P(a\neq b,b) = P(+,+) - P(-,+) = \frac{1}{4}[(1+\mu) - (1-\mu)] = \mu/2$.

and can be diagonalized analytically: its spectrum is

$$\{\lambda_1 = 0,\ \lambda_2 = r,\ \lambda_3 = 2u,\ \lambda_4 = r + 2u\}. \tag{1.17}$$

The left eigenvectors are

$$\begin{cases} u_1^T = (1, 1, 1, 1), \\ u_2^T = (-1, \frac{-u+r}{u}, \frac{+u-r}{u}, 1), \\ u_3^T = (-1, 1, -1, 1), \\ u_4^T = (1, \frac{-u-r}{u}, \frac{-u-r}{u}, 1). \end{cases} \tag{1.18}$$

The right eigenvectors are

$$\begin{aligned} v_1 &= \tfrac{1}{2(r+2u)} \begin{pmatrix} r+u \\ +u \\ +u \\ r+u \end{pmatrix}, v_2 = \tfrac{1}{2(r-2u)} \begin{pmatrix} +u \\ +u \\ -u \\ -u \end{pmatrix}, \\ v_3 &= \tfrac{1}{2(r-2u)} \begin{pmatrix} -r+u \\ -u \\ +u \\ +r-u \end{pmatrix}, v_4 = \tfrac{1}{2(r+2u)} \begin{pmatrix} +u \\ -u \\ -u \\ +u \end{pmatrix}. \end{aligned} \tag{1.19}$$

Using the expressions introduced in Appendix A.1, we find that $P(x', z)$ is given by the following 2×2 matrix:

$$P(x', z) = \begin{pmatrix} \frac{\left(1+e^{-2tu}\right)r^2+2\left(-2e^{-rt}+e^{-2ut}\right)ru-4u^2}{4(r-2u)(r+2u)} & \frac{\left(1-e^{-2ut}\right)r^2+2\left(+2e^{-rt}-e^{-2ut}\right)ru-4u^2}{4(r-2u)(r+2u)} \\ \frac{\left(1-e^{-2ut}\right)r^2+2\left(+2e^{-rt}-e^{-2ut}\right)ru-4u^2}{4(r-2u)(r+2u)} & \frac{\left(1+e^{-2ut}\right)r^2+2\left(-2e^{-rt}+e^{-2ut}\right)ru-4u^2}{4(r-2u)(r+2u)} \end{pmatrix}. \tag{1.20}$$

We can then explicitly compute $I[x_t, z_0]$, which, after some algebraic manipulation, reads:

$$\begin{aligned} I[x_t, z_0] = &\frac{1}{2}\left(1 + \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)}\right)\log_2\left[1 + \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)}\right] + \\ &+ \frac{1}{2}\left(1 - \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)}\right)\log_2\left[1 - \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)}\right]. \end{aligned} \tag{1.21}$$

If we introduce the quantity

$$\mu = \frac{-4e^{-rt}ru + e^{-2tu}r(r+2u)}{(r-2u)(r+2u)}, \tag{1.22}$$

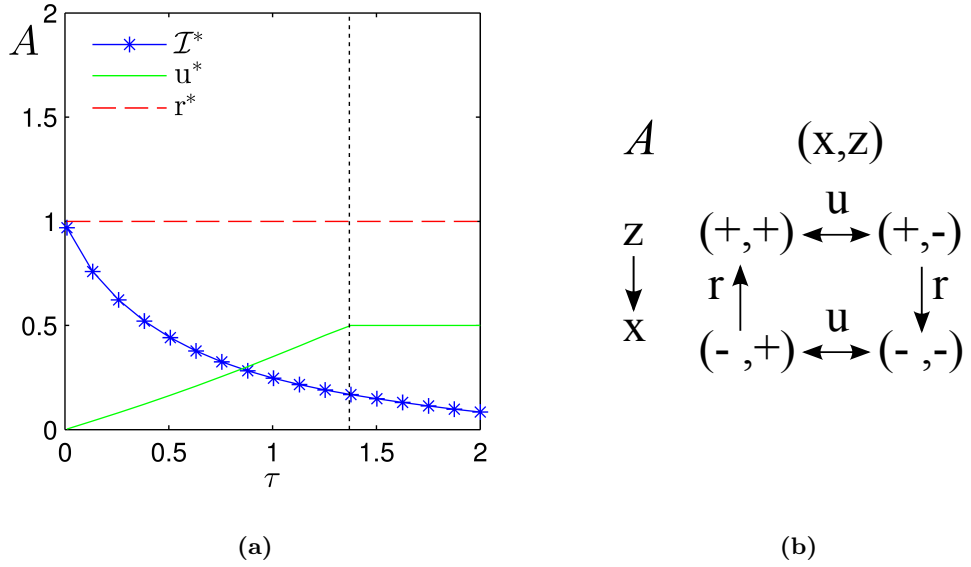
we recover the general expression for mutual information derived in Eq. 1.15. For long times, we see that $\mu \to 0$ and $I[x_t, z_0] \to 0$, as expected. Moreover, on all time-scales we find that in the stationary state

$$P(x' = z)/P(x' = -z) = (r/u + 1) > 1.$$

As in [41], we can show that the mutual information for a system initially in steady state has a maximum for a non-zero delay t^*, which is determined by the interplay of the two timescales u and r introduced above. From Eq. 1.21, taking $\frac{\delta I}{\delta t}|_{t^*} = 0$ and replacing μ, we get

$$t^* = \begin{cases} \log\left(\frac{2u+r}{2r}\right)/(2u-r), & r \neq 2u, \\ \frac{1}{2r}, & r = 2u. \end{cases} \tag{1.23}$$

However, we are not interested in finding the timescales over which information transmission is maximal, but rather the rates that maximize $\mathcal{I}(\tau)$ at *fixed* $\tau = \lambda t$, where $\lambda = \min\{2u, r\}$ in model A (optimal information curves are obtained as explained in Fig. A.1). It is worth noting that, in un-rescaled time, maximal information transmission occurs for $u \to 0$ with r finite, for any $t > 0$, as $I[x_t, z_0] \to 1$.



(a) **(b)**

Figure 1.6: Model A (activator in steady state). We show in (a) the optimal rates, along with transmitted information $\mathcal{I}^*$, as functions of the rescaled delay τ. The corresponding topology is shown in Fig. (b). Shown results are valid for nonzero delays τ and subscripts m, p are omitted when $x_m = x_p$ (with $x = u, d, r, s$).

The dynamics of the simplest model A , where z only activates x, can be summarized by the network topology shown in Fig. 1.6b. The optimal information $\mathcal{I}^*(\tau)$ and parameters $(u^*(\tau), r^*(\tau))$ are plotted in Fig. 1.6a. We see a clear crossover in terms of the switching rate u^* that regulates the state of the input z (see dashed vertical line in Fig. 1.6a): it initially increases in time and then plateaus at a value of $u^* = 0.5$. This crossover results from the fact that for $r > 2u$ the relaxation time is dominated by the rate at which the input changes ($\lambda_{\text{relaxation}} = \lambda_3 = 2u$), whereas for $r \leq 2u$ the rate at which the output changes fixes relaxation times ($\lambda_{\text{relaxation}} = \lambda_2 = r$).
Information transmission is dominated, over short timescales, by the faster rate r. Over long timescales, optimality is achieved by matching the characteristic times of the two processes, that are equal to the inverse of the two smallest non-zero eigenvalues $\lambda_2 = r$ and $\lambda_3 = 2u$. The degeneracy of the two smallest non-zero eigenvalues for large τ is a non-trivial generic feature of optimal networks that we also find in more complex models (see below).
The above behavior of the rates could also have a physical interpretation: for small delays τ, the output must react as fast as possible to the input ($r/2u \gg 1$), so as to match its state with the input state. Its optimal behavior is to react on the same timescale of the delay. However, when delay is above a certain threshold, the optimal response of the output is to react on the same timescale of the input ($r = 2u$), in order not to be too slow. Moreover, for large timescales, the input has already decorrelated from its initial state, so there is no need for the output to react extremely fast to it.

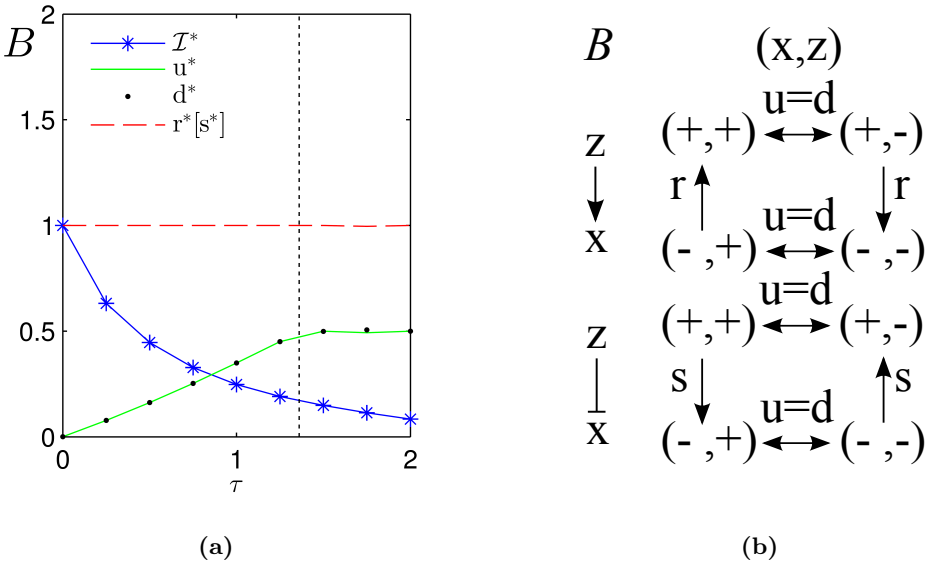
1.4.2.3 Activation and repression

We can generalize model A by allowing z to regulate x asymmetrically – that is, $r_p \neq r_m$ – and to down-regulate it as well — that is, to allow $s_m, s_p \neq 0$ (model B). As in model A , we forbid feedback from x to z, hence the transition rates for z do not depend on the state of x (i.e. $u_m = u_p \equiv u$ and $d_p = d_m \equiv d$).

Optimization yields only solutions coinciding with that of model A , or with its symmetric counterpart (wherein z perfectly down-regulates x instead of perfectly up-regulating it: $r_p = r_m = 0$ and $s_p = s_m \equiv s$). Optimal rates are shown in Fig.1.7a and the corresponding optimal topologies are depicted in Fig.1.7b.

Intuitively, in order for information to be transduced between x and z, they either align or anti-align, resulting in the common simple activator or repressor element [20]. Note that the same topological structure is found at all timescales τ. This is to be contrasted with previous studies [23, 29] which, taking into account the molecular cost paid by producing higher copy number (e.g. creating more proteins), have found small

discrepancies between the information transmitted in the two cases of up-regulation and down-regulation. Since our model does not explicitly account for protein copies, we do not observe such a difference.



(a) **(b)**

Figure 1.7: Model B (activator/repressor in steady state). We show in (a) the optimal rates, along with transmitted information $\mathcal{I}^*$, as functions of the rescaled delay τ. The corresponding topologies are shown in (b), for the activator case (top) and the repressor case (bottom). Shown results are valid for nonzero delays τ, subscripts m, p are omitted when $x_m = x_p$ (with $x = u, d, r, s$) and parameters in square brackets refer to alternative optimal topologies.

1.4.2.4 Role of feedback

Recent studies [25, 35, 36] have pointed to the important role of feedback in transmitting information, a form of which we can consider using the full set of 8 rates in Fig. 1.5 (model C). Now the hierarchical relation between z and x is broken: both can regulate each other's expression, either by down- or up-regulation.

The maximally informative topologies for all possible rescaled delays τ are illustrated in Fig. 1.8b and reveal a "push-pull" network - one gene (or protein) up-regulates the other, which in turn down-regulates the first gene/protein. Such push-pull circuits are very common in biology, from microbes [42] to humans ([43] and references therein) as a source of oscillations [46] and pulse responses [43] (see discussion below).
Due to the symmetry of the problem, we can flip either x or z and we find four equally informative solutions (shown in Fig. 1.8b), associated with different sets of the 8 rates

being driven to zero by the optimization procedure. Qualitatively, these topologies may all be described as either

1. z activates x, which in turn represses z, or
2. z represses x, which in turn activates z.

As an example, we consider a topology of the first type. Numerical optimization allows us to observe that certain rates are zero. This fact enables us to perform further analytical calculations. In particular we have that:

- $u_m = d_p$,
- $u_p = d_m = 0$,
- $r_m = r_p$,
- $s_p = s_m = 0$.

Now the transition rate matrix $\mathcal{L}$ is given by

$$\mathcal{L} = \begin{pmatrix} u_m & 0 & -r_m & 0 \\ -u_m & r_m & 0 & 0 \\ 0 & 0 & r_m & -u_m \\ 0 & -r_m & 0 & u_m \end{pmatrix} \tag{1.24}$$

and its spectrum is

$$\begin{cases} \lambda_1 = 0, \\ \lambda_2 = r_m + u_m, \\ \lambda_3 = \frac{1}{2}\left(r_m + u_m - \sqrt{r_m^2 - 6r_m u_m + u_m^2}\right), \\ \lambda_4 = \frac{1}{2}\left(r_m + u_m + \sqrt{r_m^2 - 6r_m u_m + u_m^2}\right). \end{cases} \tag{1.25}$$

For any value of rates u_m and r_m, the smallest nonzero eigenvalue is always λ_3. The left eigenvectors are

$$\begin{cases} u_1^T = (1, 1, 1, 1), \\ u_2^T = (1, -\frac{r_m}{u_m}, -\frac{r_m}{u_m}, 1), \\ u_3^T = (-1, \frac{+r_m - u_m - \sqrt{r_m^2 - 6r_m u_m + u_m^2}}{2r_m}, \frac{-r_m + u_m + \sqrt{r_m^2 - 6r_m u_m + u_m^2}}{2r_m}, 1), \\ u_4^T = (-1, \frac{+r_m - u_m + \sqrt{r_m^2 - 6r_m u_m + u_m^2}}{2r_m}, \frac{-r_m + u_m - \sqrt{r_m^2 - 6r_m u_m + u_m^2}}{2r_m}, 1). \end{cases} \tag{1.26}$$

The right eigenvectors are

$$
\begin{aligned}
v_1 &= \frac{1}{2(r_m+u_m)}\begin{pmatrix} r_m \\ u_m \\ u_m \\ r_m \end{pmatrix}, v_2 = \frac{1}{2(r_m+u_m)}\begin{pmatrix} +u_m \\ -u_m \\ -u_m \\ +u_m \end{pmatrix}, \\
v_3 &= \begin{pmatrix} \frac{-r_m+u_m-\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{-u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{+u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{+r_m-u_m+\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \end{pmatrix}, v_4 = \begin{pmatrix} \frac{+r_m-u_m-\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{+u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{-u_m}{2\sqrt{r_m^2-6r_mu_m+u_m^2}} \\ \frac{-r_m+u_m+\sqrt{r_m^2-6r_mu_m+u_m^2}}{4\sqrt{r_m^2-6r_mu_m+u_m^2}} \end{pmatrix}
\end{aligned} \tag{1.27}
$$

We can again express the mutual information as in Eq. 1.15, with μ given by

$$
\begin{aligned}
\mu &= \cosh(t\sqrt{u_m^2+r_m^2-6r_mu_m})\frac{r_m-u_m}{r_m+u_m} + \\
&+ \sinh(t\sqrt{u_m^2+r_m^2-6r_mu_m})\frac{u_m^2-r_m^2-4r_mu_m}{(r_m+u_m)\sqrt{u_m^2+r_m^2-6r_mu_m}}.
\end{aligned} \tag{1.28}
$$

For $t=0$, $\mu_0 = \dfrac{r_m-u_m}{r_m+u_m}$. For large t we can define

$$x = t\sqrt{u_m^2+r_m^2-6r_mu_m}$$

and, noting that

$$\lim_{x\to 0}\frac{\sinh x}{x} = 1,$$

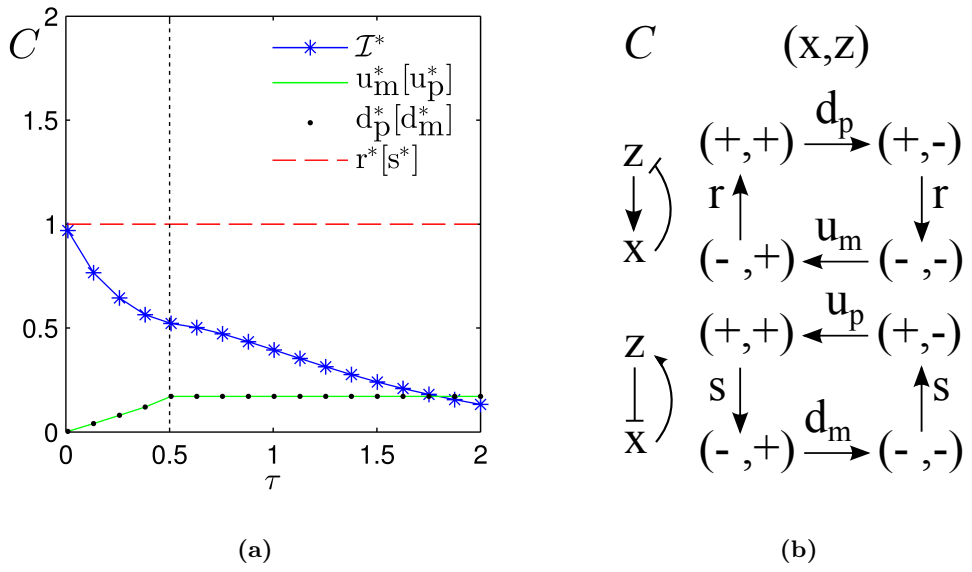
we obtain

$$\mu = \frac{r_m-u_m}{r_m+u_m} + \frac{u_m^2-r_m^2-4r_mu_m}{r_m+u_m}t.$$

As for models A and B , we are interested in finding the rates that maximize $\mathcal{I}(\tau)$ at fixed $\tau = \lambda t$, where λ is the smallest nonzero eigenvalue (λ_3 for the example worked out in Eq.1.25).

The numerical results are shown in Fig. 1.8a. We find that the optimal value of the input rates $(u_m[u_p], d_p[d_m])$ now plateau at a value of $3-2\sqrt{2}$ for $\tau > \tau^* = 0.5$ (see dashed vertical line): this value is set by competition with the $r[s]$ rate, by matching the smallest non-zero eigenvalue λ_3 with λ_4 in order to avoid oscillatory solutions (for input rates $> 3-2\sqrt{2}$, the above eigenvalues become imaginary describing oscillations).

Push-pull networks can oscillate [46] [3], thwarting optimal information transmission by decorrelating the system, hence the oscillatory regime is not the optimal solution.



(a) **(b)**

Figure 1.8: Model C (activator/repressor with feedback in steady state). We show in (a) the optimal rates, along with transmitted information $\mathcal{I}^*$, as functions of the rescaled delay τ. The corresponding topologies are shown in (b), for the activator case (top) and the repressor case (bottom): they both represent a "push-pull" network, where one gene (or protein) activates the other, which in turn represses the first gene/protein (as we can see, the roles of z and x are interchangeable). Shown results are valid for nonzero delays τ, subscripts m, p are omitted when $x_m = x_p$ (with $x = u, d, r, s$) and parameters in square brackets refer to alternative optimal topologies.

The optimal solution exploits feedback to transmit more information. For small delay times feedback does not play a role and model C transmit the same amount of information as models A and B . For delay times at which the input rates have reached their plateau value, the optimal circuit of model C can transmit more information at a fixed τ than the optimal circuits without feedback (compare Fig. 1.8a with Figs. 1.6a and 1.7a). Intuitively, this happens because the value of the switching rates of z at the plateau is smaller in model C than in models A and B , thus z is less likely to switch on/off. As a result of these slower switching rates, the system in models A and B is more likely to cycle through the four states and hence to obscure correlations with the initial condition than in model C . Additionally, feedback leads to a rotational directionality among the transitions that is not observed in simpler models (cf. Fig. 1.8b). As a result of this rotational directionality the system never directly 'flips back', enhancing the transduced information. In summary, feedback allows for a combination of slower flipping rates and

[3] Delay in mRNA production was shown to be a necessary element of stable oscillations, making them hard to observe in synthetic networks [46].

imposed order to the visited states that enables to read out more information about the initial state at later times.

1.4.2.5 Systems out of steady state

Having discussed optimal delayed information transmission of repeated readouts in the stationary state, we now consider regulatory networks that optimize a one-time response to an input (e.g. by producing an enzyme to metabolize a nutrient that appeared in the environment).

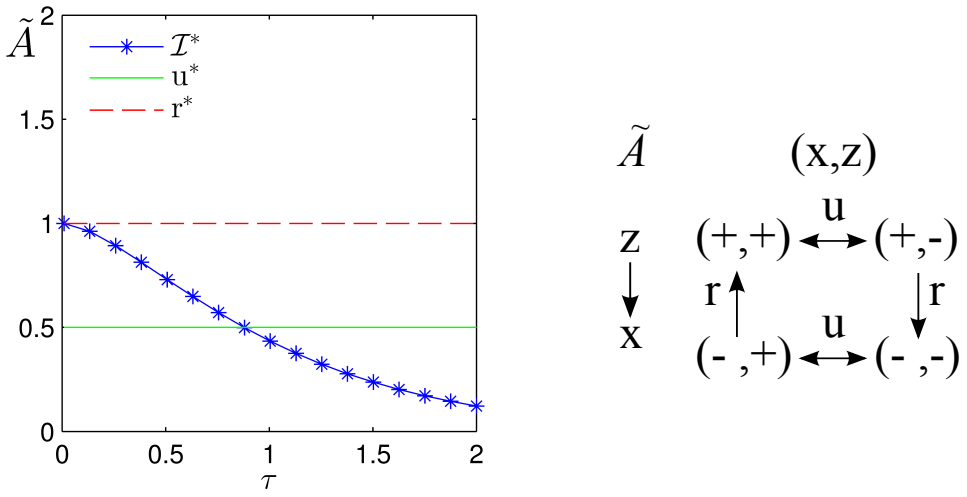
We assume that at time t_0, the system is in an initial state which is not its steady state. We ask what is the optimal design of the circuit to produce a most informative output to this initial state given a fixed delay. Unlike in the previous case where the network was in steady state (which determined the initial distribution) and would respond repeatedly, we now also ask what is the optimal initial distribution of the system. We allow the initial and final distribution to be different. The optimization over the input distribution describes the matching of the properties of the regulatory network and the initial non-stationary state corresponding, e.g., to the appearance of a sugar source in the cell, DNA damage or food shortages. Specifically, we consider the same three models studied above (A , B , C), but now optimize simultaneously not only on the rates but also on the initial probability distribution $P_0(y)$ and refer to the associated models as $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$, respectively.

To calculate the capacity of the system we optimize over the initial distribution and the parameters of the network. We could fix the initial probability distribution and optimize only over the network parameters. However this would be an arbitrary choice of the initial distribution and we would not calculate the capacity of the system. Instead we consider all possible input distributions and ask which one of them guarantees maximal information transmission. Cells often pre-process external signals, for example the lack of glucose is presented in terms of high cAMP (or more specifically activated crp) concentration to the lac operon, which gives the cell a certain degree of control over the distribution of the input to a network.

We start with model $\tilde{A}$, which enjoys the usual symmetry $-1 \leftrightarrow +1$ for x and z. This constrains the form of the initial distribution to be parameterized as

$$P_0(x,z) = \left(\frac{1+\mu_0}{4}, \frac{1-\mu_0}{4}, \frac{1-\mu_0}{4}, \frac{1+\mu_0}{4}\right) \text{ [4]}. \tag{1.29}$$

[4]The case of a steady state initial distribution is a special case of a system that enjoys this property.



(a) **(b)**

Figure 1.9: Model $\tilde{A}$ (activator out of steady state). We show in (a) the optimal rates, along with transmitted information $\mathcal{I}^*$, as functions of the rescaled delay τ. The corresponding topology is shown in (b). The optimal initial distribution $P_0(y)$, which is not displayed, concentrates on states $(+,+)$ and $(-,-)$ at all times. Shown results are valid for nonzero delays τ,subscripts m, p are omitted when $x_m = x_p$ (with $x = u, d, r, s$).

The probability $P(x', z)$ then reads

$$P(x',z) = \begin{pmatrix} \frac{r+e^{-2tu}r-2u+e^{-rt}(-r+r\mu_0-2u\mu_0)}{4(r-2u)} & \frac{r-e^{-2tu}r-2u+e^{-rt}(+r-r\mu_0+2u\mu_0)}{4(r-2u)} \\ \frac{r-e^{-2tu}r-2u+e^{-rt}(+r-r\mu_0+2u\mu_0)}{4(r-2u)} & \frac{r+e^{-2tu}r-2u+e^{-rt}(-r+r\mu_0-2u\mu_0)}{4(r-2u)} \end{pmatrix}, \quad (1.30)$$

and we can explicitly compute the mutual information:

$$I[x_t, z_0] = \frac{1}{2}\left(1+\mu_0 e^{-rt}+\frac{r}{r-2u}(e^{-2ut}-e^{-rt})\right)\log_2\left[1+\mu_0 e^{-rt}+\frac{r}{r-2u}(e^{-2ut}-e^{-rt})\right]+ \\ + \frac{1}{2}\left(1-\mu_0 e^{-rt}+\frac{r}{r-2u}(e^{-2ut}-e^{-rt})\right)\log_2\left[1-\mu_0 e^{-rt}+\frac{r}{r-2u}(e^{-2ut}-e^{-rt})\right]. \quad (1.31)$$

The above expression for the mutual information can again be rewritten as in Eq.1.15 if we introduce

$$\mu = \mu_0 e^{-rt} + \frac{r}{r-2u}(e^{-2ut}-e^{-rt}). \quad (1.32)$$

For $t = 0$, $\mu = \mu_0$ and the information is maximized by $\mu = \mu_0 = \pm 1$, that is when z and x are perfectly aligned/anti-aligned. Moreover, μ and $I[x_t, z_0]$ decrease exponentially with time t. Therefore, unlike in model A , the information transmission does not

improve by making a delayed readout. In other terms, the absence of a maximum at $t^* > 0$ in $I[x_t, z_0]$ for optimal initial states suggests that, at odds with the stationary case (model A), the mechanism for information transmission is only governed by the loss of information about the initial state as the system relaxes to stationarity.

After performing the optimization of $\mathcal{I}(\tau)$, we find that for each rescaled delay τ the optimal initial distribution $P_0(y)$ concentrates on the states $(+,+)$ and $(-,-)$. We can understand this result intuitively: the rate for switching out of these states, u, is small, so the system is more likely to remain in these states than in the other two states (see Fig. 1.9a and Fig. 1.9b). Posing the system in these long-lived states allows for more information transmission about the initial condition at small readout delays τ.

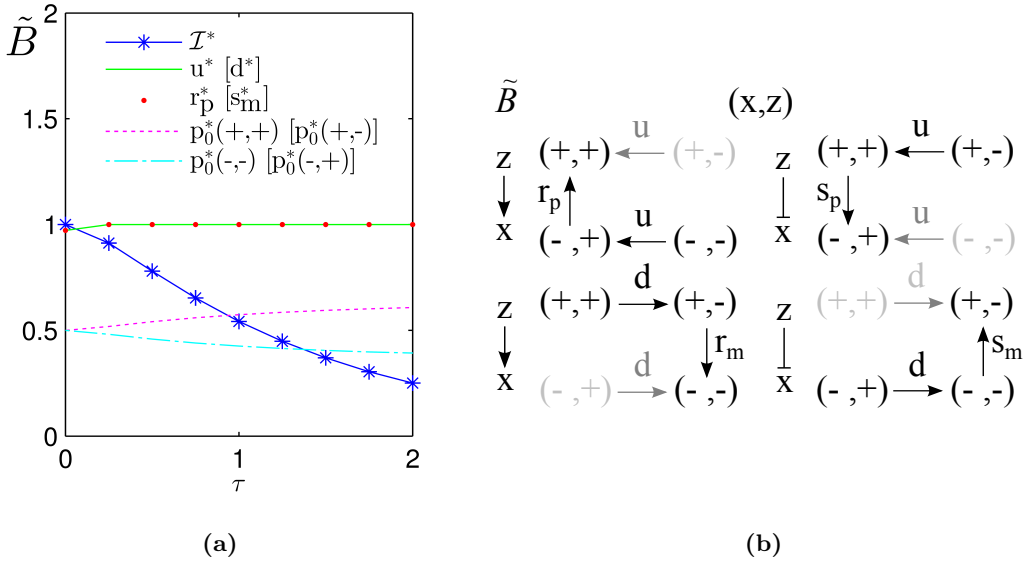


Figure 1.10: Model $\tilde{B}$ (activator/repressor out of steady state). We show in (a) the optimal rates and initial distribution, along with transmitted information $\mathcal{I}^*$, as functions of the rescaled delay τ. The corresponding topologies are shown in (b), for the activator case (left) and the repressor case (right): each topology features a different absorbing state. States displayed in gray are never visited. Shown results are valid for nonzero delays τ, subscripts m, p are omitted when $x_m = x_p$ (with $x = u, d, r, s$) and parameters in square brackets refer to alternative optimal topologies.

We now turn to maximizing information transmission over the initial distribution $P_0(y)$ in the more general models $\tilde{B}$ and $\tilde{C}$. As above, symmetry provides a number of optimal networks related by permutations (see Fig. 1.10b and Fig. 1.11b). However, unlike in model $\tilde{C}$ where all states are visited, in each optimal setting of model $\tilde{B}$ one state (gray in Fig. 1.10b) is never visited. The optimal rates are shown for the case of z up-regulating [down-regulating] x in Fig. 1.10a and Fig. 1.11a.

We find a qualitative difference in design as compared to the stationary case (models

B and C): while the optimal topology remains the same, now either one of the aligned or non-aligned states becomes **absorbing** [5].

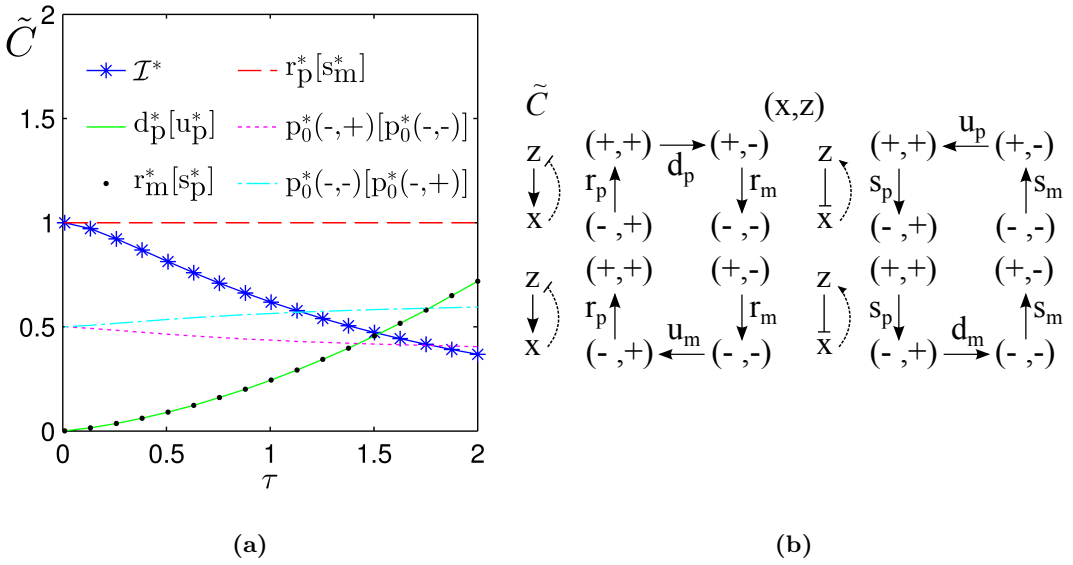


Figure 1.11: Model $\tilde{C}$ (activator/repressor with feedback out of steady state). We show in (a) the optimal rates and initial distribution, along with transmitted information $\mathcal{I}^*$, as functions of the rescaled delay τ. The corresponding topologies are shown in (b), for the activator case (left) and the repressor case (right): they all represent a "push-pull" network, where one gene (or protein) activates the other, which in turn represses the first gene/protein (as we can see, the roles of z and x are interchangeable). Moreover, each topology features a different absorbing state (dashed lines mean that feedback exists only until the absorbing state is reached). Shown results are valid for nonzero delays τ, subscripts m, p are omitted when $x_m = x_p$ (with $x = u, d, r, s$) and parameters in square brackets refer to alternative optimal topologies.

The occurrence of an absorbing state, with a nearly-equal optimal initial distribution $P_0(y)$ over the initial and final states, limits the system's dynamics and leads to the optimal topology for a one time response. In the absence of feedback (model $\tilde{B}$, e.g. receptor activation in a complex pathway), when the system, initially in the inactive state $(x, z) = (-, -)$ is presented with a signal $(x, z) \to (-, +)$, it switches on a response $(x, z) \to (+, +)$ (see Fig. 1.10b). However, in the presence of feedback (model $\tilde{C}$, e.g. a nutrient activating the production of an enzyme for its uptake, amino acid biosynthesis) the optimal dynamics includes "feedback inhibition", in which the output switches off the input (see Fig. 1.11b) [37] [6]. As in model C , feedback imposes an order to the visited

[5] In each of the four degenerate topologies, a different state becomes absorbing.

[6] Such design is reminiscent of Shannon's Ultimate Machine: a device with a switch that, when it is turned on, activates an arm that switches it off.

states, with a smaller rate for z transitions than for x transitions: these two features allow again for higher information transmission about the initial state (see Fig. 1.11a).

Let us consider a specific biological example - lactose metabolism - for the optimal network shown in Fig. 1.11b (top left panel) and its corresponding optimal rates and initial distribution presented in Fig. 1.11a. The system describes two elements: lactose, z, and the degrading enzyme beta-galactosidase in the lac operon, x. Our optimal solution consists of the rates and topologies of the network and the initial probability distribution of the system. The optimal solution is based on the matching between the statistics of the input and the output (lactose and beta-galactosidase) and the properties of the network, similarly to the approach taken in neuroscience by Laughlin [47] and in studying time-independent models of information transmission in molecular systems [23–25, 29]. We find that the optimal initial distribution limits the system to be with close to equal probability in either the $(x, z) = (-, +)$ or the $(x, z) = (-, -)$ state, which correspond to states in which the enzyme is turned off and lactose is either there or not. These initial states are implemented using the lac repressor, that senses whether lactose is present or not in the environment and represses the lac operon when needed.

Our calculation does not fix the initial state of network, but assigns initial probabilities to all states. This network design is optimal given this nearly equal probability of the input sensing sugar in the environment (for example, activated crp in response to glucose in the lac operon). However, we did not constrain the initial distribution, but we asked for a best matching between the properties of the circuit and any initial distribution. If the probability distribution of the sugar in the environment were known and fixed, we would have to optimize the network given this additional constraints. The calculation presented in models A , B and C is a specific example of when the input probability distribution is constrained to be the steady state distribution of the system (and therefore uniquely set by the optimal rates).

According to our optimal solution, lactose can either be present or not in the cell. In both cases the degradation enzymes are switched off. If there is no lactose, the system at t_0 is in the final absorbing state $(x, z) = (-, -)$. If at t_0 lactose (z) is sensed, the enzymes x that degrade sugar are switched off and the system is in the $(x, z) = (-, +)$ state. The appearance of lactose activates the synthesis of enzymes $((-, +) \to (+, +))$, which cause the depletion of the sugar $((+, +) \to (+, -))$. Finally the lack of lactose deactivates the enzymes and they enzymes are degraded $((+, -) \to (-, -))$. This matching between the initial probability distribution of seeing sugar in the environment and the regulatory elements of the network allows the system to transmit most information about the original state of the input with a delay. Given the fast initial rate for leaving the

state $(x, z) = (-, +)$, if the enzyme is present, then sugar was initially present. If there is no enzyme, there was no sugar [7].

The signal can directly be the input, as in the lactose metabolism example, or it can influence the input. A biological example of the push-pull network shown in Fig. 1.11b (top left panel) with the external signal that triggers an input is the p53-MdM2 circuit that is involved in DNA repair [43, 48]. The tumor suppressor protein p53 transcriptionally activates the MdM2 gene, the product of which degrades p53. DNA damage leads to an increase in p53 levels ($(x, z) = (-, +)$), which in turn up-regulates MdM2 ($(-, +) \rightarrow (+, +)$) that degrades p53 ($(+, +) \rightarrow (+, -)$) and in turn down-regulates MdM2 ($(+, -) \rightarrow (-, -)$). In our optimal solution the initial distribution of this network is tuned to a roughly equal probability of there being DNA damage or not.
In both the lactose and DNA damage case the optimal networks perform one readout, after which they need to be reset externally: additional sugar needs to be taken up from the environment or p53 levels need to be increased by new DNA damage.

1.4.3 Discussion and conclusions

In Fig. 1.12 we plot a comparison of all the cases possible within the two-state model. As the model generality increases (from A to $\tilde{C}$), so does the number of parameters; accordingly, the information capacity of the system also increases. As explained in section 1.4.2.4, we see that the introduction of feedback in model C does not play a role in increasing information transmission for small τ. However, the information gain coming from feedback is substantial for long rescaled delays τ between the input and output readout. Information transmission can be improved beyond that achieved in the steady state (A , B and C) if the system is pushed out of equilibrium in specific ways ($\tilde{A}$, $\tilde{B}$ and $\tilde{C}$) to respond to one time signals (as explained in subsection 1.4.2.5). This increase in information is achieved by simultaneously optimizing the initial distribution of inputs and outputs in a way that matches the properties of the regulatory circuit.

Both in the steady state and the non-steady state feedback models we find the optimal network topology of a push-pull circuit. Such circuits exist in many cells, ranging from bacterial (heat-shock response [42]) to mammalian (I-κB-NFκB circuit [44] in animal stress response, p53-Mdm2 network involved in DNA damage response [43]), and often combine a slow (transcriptional regulation) with a fast (protein-protein interaction) component, similarly to the design of our optimal architectures. In particular, the non-steady state optimal topologies feature absorbing states, which result in single pulse responses. These responses are common in the case of stress signals and in some cases

[7] In a real system the enzymes could be present at a basal level even with the absence of sugar, but our coarse-grained model cannot account for that.

feature a "digital" behavior: the number of pulses, rather than the response intensity, is proportional to the input strength [43].

We find that allowing for feedback (models C and $\tilde{C}$) results in optimal solutions that transmit more information than models without feedback. This observation is similar to those found for instantaneous information transmission [24] and for the rate of information transmission [35]. Our optimal solutions consist of two elements: the combination of slow and fast rates of reactions and the imposed rotational order of the states. In non-equilibrium circuits where detailed balanced can be broken, the order of visited states is further enforced by the appearance of an absorbing state, which means that each state can only be visited once. These two general design elements make the readout of the input from the output more distinguishable, which in turn increases the amount of transmitted information, similarly to what was previously found for more detailed molecular models and instantaneous information transmission. As noted, many push-pull networks that take part in stress response have a slow and a fast timescale. Other molecular implementations of these general principles may be possible, however this simple model points to very general design elements.

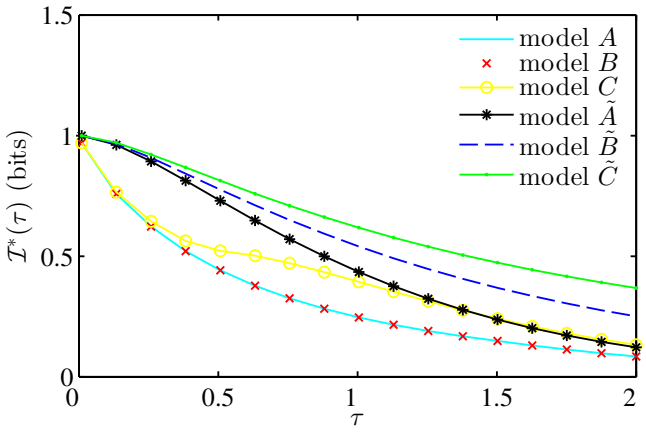


Figure 1.12: Summary of optimization results: for each of the six models, the hull curve of the maximized information $\mathcal{I}(\tau)$ is plotted versus τ. When feedback is present (model C) information is higher for $\tau \gtrsim 0.4$. When the system is initially in an optimal state (model $\tilde{A}$, $\tilde{B}$, $\tilde{C}$) the information is higher for each τ and its time decay is qualitatively different.

Chapter 2

The energetic cost of information transmission

2.1 Stochastic thermodynamics: a short review

Nonequilibrium systems with discrete states can be described [49–51] by a master equation of the form

$$\frac{dP_i(t)}{dt} = \sum_{j=1}^{N} -w_{ij}P_i(t) + w_{ji}P_j(t), \tag{2.1}$$

where w_{ij} is the transition rate from state i to state j and $P_i(t)$ denotes the probability of state i at time t, which satisfies

$$0 \leq P_i \leq 1, \quad i = 1, \ldots, N, \quad \sum_{i=1}^{N} P_i(t) = 1.$$

One can define the probability current J_{ij} as

$$J_{ij} = -w_{ij}P_i + w_{ji}P_j, \tag{2.2}$$

which expresses the difference in the frequency of transitions from state i to j and backward per unit time.

The above definition for J_{ij} implies that $J_{ij} = -J_{ji}$ and we can easily assume that $J_{ii} = 0$ for all i. Then we can rewrite the master equation 2.1 in the following way:

$$\frac{dP_i(t)}{dt} = \sum_{j=1}^{N} J_{ij}. \tag{2.3}$$

If there exists a stationary solution $P_i = \lim_{t\to\infty} P_i(t)$ to the above equation, then the probability currents J_{ij} satisfy

$$\sum_{j=1}^{N} J_{ij} = 0 \quad \forall i, \tag{2.4}$$

which reminds of the Kirchhoff's Current Law for each state i.
The condition of **detailed balance**

$$J_{ij} = -w_{ij}P_i + w_{ji}P_j = 0 \quad \forall i, j \tag{2.5}$$

is sufficient but unnecessary for the existence of a stationary solution. Indeed a stationary state can exist also in the presence of probability currents between the states, if they form a suitable current loop structure.
The nonvanishing probability currents in the stationary state of a nonequilibrium system are related to the total entropy production rate, which is defined as

$$\sigma_{\text{tot}} = \sum_{ij} P_i(t) w_{ij} \log \frac{P_i(t) w_{ij}}{P_j(t) w_{ji}}. \tag{2.6}$$

If we use the probability currents J_{ij} defined above and we introduce the *affinity* $A_{ij} = \log \frac{P_i(t)w_{ij}}{P_j(t)w_{ji}}$, we can rewrite σ_{tot} as

$$\sigma_{\text{tot}} = \frac{1}{2} \sum_{ij} J_{ij} A_{ij}. \tag{2.7}$$

We observe that the entropy production rate of Eq. 2.6 can be split in two contributions $\sigma_{\text{tot}} = \sigma_{\text{eq}} + \sigma$, where

$$\sigma_{\text{eq}} = \sum_{ij} P_i(t) w_{ij} \log \frac{P_i(t)}{P_j(t)} \tag{2.8}$$

and

$$\sigma = \sum_{ij} P_i(t) w_{ij} \log \frac{w_{ij}}{w_{ji}}. \tag{2.9}$$

By using the master equation 2.1, one can easily show that $\sigma_{\text{eq}} = \frac{dS}{dt}$, where S is system's entropy under **equilibrium** conditions:

$$S = -\sum_{i} P_i \log P_i. \tag{2.10}$$

We can then conclude that the contribution σ is due to the coupling of the system to external thermodynamic forces which prevent it from reaching equilibrium and can be interpreted as the rate of entropy increase in the external medium. If there was also an external dissipative process, the total entropy production σ would have to include another term σ_{ext} to account for the external production of entropy.

2.2 Modeling the trade-off between information and energy dissipation

Cells respond to the current state of their environment by processing external signals through molecular networks and cascades. An external chemical stimulus is measured by receptors, which activate a series of biochemical reactions and lead the cell to produce an appropriate response. This response can be activating a gene or pathway, producing proteins that process the signal as in the case of sugar metabolism, result in motion such as in the case of chemotaxis, or initiating a cellular response such as apoptosis. As we learn more about the structure of biochemical networks, we need to understand the functional role of their elements and connections. Yet regulation comes at a cost, which imposes constraints on the form of these networks. Here we consider the limitations coming from thermodynamic constraints, caused by the cell's energy consumption, on the architecture of regulatory elements that best convey information about input signals to their outputs.

Despite the large complexity of biological regulatory networks, not all possible molecular regulatory circuits can be found in living organisms [52]. One can ask whether the network architectures and parameter regimes are only shaped by the evolutionary history of these organisms, or whether there are also physical limits that constrain them. In the last years, a number of groups have explored different physical principles that could influence the parameter regimes and modes of regulation in living organisms (e.g. [2, 6–9, 11, 24, 33–35, 53–62]). One approach has been to calculate the limits that the intrinsic randomness in gene regulation imposes on information transmission between the input signal and its output responses, in networks of varying complexity [23–25, 29, 30, 32, 63, 64]. These studies showed which network architectures are optimal for information transmission and found that distinguishing different output states in general increases the transmitted information. They also pointed to the important trade offs between the information that the output has about the input and molecular costs.

Many of the current approaches to information transmission have looked at instantaneous information transmission [23–25], or the rate of information transmission [33–36]. However, it has been argued that information transmission may be enhanced by dynamic biochemical readouts at multiple time points [27] or when the regulatory response is at a delay relative to input signaling [41]. Additionally, many biochemical networks function out of steady state, responding to inputs that are changing in time. Examples include the chemotactic response of bacteria or amoebas to nutrients or conversely to antibiotics.

Inspired by these observations, in Chap.1 we previously studied the optimal circuits for transmitting information between an input and output read out with a fixed delay, in and

out of steady state. Delayed readouts are natural to most biochemical circuits, since sensing a signal requires production of the response, which takes time. For example, sensing an increased sugar concentration means the cell has to produce the enzyme to degrade it. We asked whether different readout delays correspond to different optimal circuits. We found that topologies of maximally informative networks correspond to commonly occurring negative feedback circuits irrespective of the temporal delay specified. Most interestingly, circuits functioning out of steady state may exploit non-equilibrium absorbing states to transmit information optimally and feedback can additionally increase information transmission. We found that there are many degenerate topologies that transmit similar information equally optimally - a degeneracy that will most likely be lifted by considering more detailed molecular models.

The optimal solutions we found previously function strongly out of equilibrium, so they must consume energy. Since it has been experimentally shown [65] that sensory systems may have evolved to reduce their energy expenditure, we were interested in seeing how energetic constraints impact the form of the optimally informative solutions. This knowledge will prove useful when constructing artificial biochemical circuits [66], or engineering living organisms for energy production [67]. The energy dissipated (or consumed) by a given network can be estimated by looking at the thermodynamics of its composite biochemical reactions. A completely reversible reaction does not consume energy. The reaction is in perfect equilibrium and the total free energy of the system is completely balanced. Irreversible reactions, such as certain steps of biochemical cascades, come at a cost to the cell, which has to prevent the back reaction from occurring. This cost can be estimated considering the flux balance of the network. The heat dissipated by the circuit is proportional to its rate of entropy production [68]. Tu et al. [69] looked at entropy production in biochemical regulatory networks and experimentally showed that the flagellar motor switch of *Escherichia coli* operates out of equilibrium, dissipating energy. A nonequlibrium allosteric model consistent with experimental results was proposed to explain how the switch operates with high sensitivity at a small energetic cost.

Energetic cost has also been discussed in relation to cellular precision and the predictive power of the cell. The chemosensory system of *E. coli* has been shown to dissipate energy in order to improve its adaptive speed and accuracy [70]. Reliable readout of input concentrations has also been bound by the entropy production rate [71–75]. Others have reversed the perspective and shown that the minimum energy required for a biological sensor to detect a change in an environmental signal is proportional to the amount of information processed during the process [76]. In the case of the *E. coli* chemosensory system, it was argued that 5% of the energy consumed in sensing is determined by information-thermodynamic bounds, and is thus unavoidable [76]. Becker et al. [77]

showed that short-term prediction in a sensory module is possible in equilibrium, but only up to a finite time interval. For longer times accurate prediction requires large dissipation. Lastly, the inability of systems to use all knowledge of past environmental fluctuations to predict the future state has been directly linked to dissipation [78].

We want to see how the structure of optimal networks for information transmission changes if we impose a penalty on the entropy production of the system. In order to investigate the non-equilibrium nature of biochemical circuits that are optimal for delayed information transmission, we choose to study a simple binary model of a regulatory circuit that allows us to focus on the regulatory logic at small computational costs. Within this model we consider two interacting elements of biochemical regulatory networks (e.g. proteins and genes, elements of two component signaling systems, sugars and enzymes) that take on binary states (on or off) and evolve in continuous time. This simplification allows us to develop an efficient formalism for calculating information transmission at different readout delays and consider the connection between dissipation and different readout times. In the limit of infinite dissipation rates, we recover the results obtained in Chap.1. For finite, non zero dissipation rates, back reactions decrease the information transmission until it goes to zero for systems close to equilibrium. However, when feedback is allowed, networks are able to transmit almost 1 bit of information at no cost.

2.3 A model regulatory circuit for time-dependent information transmission with energy dissipation

2.3.1 Model setup

To focus on the tradeoffs between the ability of the network to transmit information and the energetic cost of the biochemical reactions that make up this network, we study the simplest model of regulation that allows us to focus on the logic of the reactions. The setup is the same as in Chap.1, nevertheless we will shortly review it hereafter.

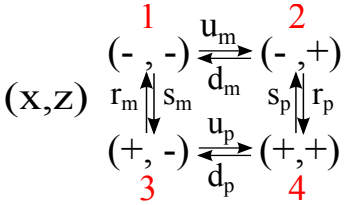


Figure 2.1: Sketch of the four network states, with corresponding transition rates. The dynamics is the same as in Fig.1.5.

We have a simplified network consisting of two binary elements (see Fig. 2.1) that describe either a transcription factor protein regulating a gene, or a signaling molecule

activating/downregulating an enzyme or receptor. The first element of the network describes the input z and can be associated with the state of a receptor, signaling molecule or transcription factor that responds to the external conditions. For example, it can describe the presence or absence of a sugar source in metabolism or phosphorylation of the histidine kinase in a two-component signaling system. The output x describes the final outcome of the network, such as the gene that produces the response protein to the external signal. In the examples given above, it corresponds to the enzyme that digests the sugar or to the expression of the target gene by the response regulator. Both of these elements can be found in the active $(x, z = +1)$ or inactive $(x, z = -1)$ states. If the described element is a continuous variable (e.g. protein concentration), the binary approximation is equivalent to taking very steep regulatory functions, such that the concentration is well described by two states: below and above the threshold.

Our two-component network can be found in one of four states:

$$(x, z) \in \{(-,-), (-,+), (+,-), (+,+)\}, \tag{2.11}$$

corresponding to both elements inactive, the input active — output inactive and vice versa, and both elements active. The input z up/down regulates x with rates $r_m(r_p)$ and $s_m(s_p)$, that depend on the state of the input $(m = -, p = +)$. The system is described by the conditional probability $P(x_t, z_t, t|x_0, z_0, 0)$ of being in state (x_t, z_t) at time t, given the state (x_0, z_0) at time $t = 0$. This conditional probability distribution can be arranged in a 4×4 matrix, and its evolution is described by the master equation

$$\partial_t P = -\mathcal{L}P \tag{2.12}$$

where the 4×4 transition matrix $\mathcal{L}$ is defined in terms of the rates depicted in Fig. 2.1, as in Eq.1.10 of Chap.1. The central quantity we shall be interested in is the joint probability distribution of the state x_t of the output at time t and the state z_0 of the input at time 0. We shall use the shorthand

$$P(x_t, z_0) = \sum_{z_t, x_0 = \pm 1} P(x_t, z_t, t|x_0, z_0, 0) P_0(x_0, z_0) \tag{2.13}$$

where $P_0(x_0, z_0)$ is the probability distribution of the system at the initial time.

In analogy with Chap.1, we are interested in finding the network topologies that are optimal for information transmission over a fixed time scale τ. Specifically we want to maximize the mutual information between an input signal at an initial time, z_0, and the output of the network which is read out at a later time, x_t. We define it as

$$\mathcal{I}(\tau) = I[x_{t=\tau/\lambda}, z_0], \tag{2.14}$$

where

$$I[x_t, z_0] = \sum_{x_t, z_0} P(x_t, z_0) \log \frac{P(x_t, z_0)}{P(x_t) P_0(z_0)}, \tag{2.15}$$

and $P_0(z_0) = \sum_{x_0} P_0(x_0, z_0)$, $P(x_t) = \sum_{z_0} P(x_t, z_0)$.
The maximization of $\mathcal{I}(\tau)$ is performed over the rates of the biochemical reactions $\mathcal{L}$ of the regulatory network. We measure the time $t = \tau/\lambda$ between the signal and the delayed read-out in units of the natural timescale of the problem – the relaxation time λ^{-1}, calculated as the inverse of the minimal non-zero eigenvalue of $\mathcal{L}$.

Previously we found the networks that are best suited for transmitting information at a delay and discovered that they correspond to systems that function out of equilibrium. For this reason we are interested in posing the same question, but taking into account energy constraints. We thus constrain the energy $\dot{Q}$ dissipated per unit time into an external medium at temperature T that is in contact with our system. $\dot{Q}$ is related to the thermodynamic entropy production rate σ (defined in Eq.2.9):

$$\dot{Q} = k_B T \sigma, \tag{2.16}$$

where k_B is the Boltzmann constant [49, 68].

Calculation of the entropy production rate

We perform the calculation of the entropy production rate σ, for the general dynamic system described by the transition rate matrix $\mathcal{L}$ and pictured in Fig.2.1.
We start from the definition of σ:

$$\sigma = \sum_{i,j} P_i w_{ij} \log \frac{w_{ij}}{w_{ji}},$$

where P_i is the stationary probability distribution P_∞ for state i. In our specific case, we explicitly have

$$\begin{aligned} \sigma &= P_1 w_{12} \log \frac{w_{12}}{w_{21}} + P_2 w_{24} \log \frac{w_{24}}{w_{42}} + \\ &+ P_4 w_{43} \log \frac{w_{43}}{w_{34}} + P_3 w_{31} \log \frac{w_{31}}{w_{13}} + \\ &+ P_2 w_{21} \log \frac{w_{21}}{w_{12}} + P_4 w_{42} \log \frac{w_{42}}{w_{24}} + \\ &+ P_3 w_{34} \log \frac{w_{34}}{w_{43}} + P_1 w_{13} \log \frac{w_{13}}{w_{31}}. \end{aligned}$$

After collecting similar terms we can write

$$\begin{aligned}\sigma &= J_{12}\log\frac{w_{12}}{w_{21}} + J_{24}\log\frac{w_{24}}{w_{42}} + \\ &+ J_{43}\log\frac{w_{43}}{w_{34}} + J_{31}\log\frac{w_{31}}{w_{13}},\end{aligned}$$

where we have used the definition of probability current J_{ij}, introduced in Eq.2.2, and we have considered a CW cycle in Fig.2.1. One can easily show that all the currents are equal to each other:

$$J_{12} = J_{24} = J_{43} = J_{31} = J. \tag{2.17}$$

Hence the entropy production rate σ can be simply cast as

$$\sigma = J\log\frac{w_{12}\,w_{24}\,w_{43}\,w_{31}}{w_{21}\,w_{42}\,w_{34}\,w_{13}}. \tag{2.18}$$

We then plug in the above expression the rates of the transition matrix $\mathcal{L}$ and the stationary distribution P_∞. In our specific case, P_∞ is equal to the right eigenvector v_1 of $\mathcal{L}$, which corresponds to the null eigenvalue λ_1 and is given by

$$v_1 = \frac{1}{\sum_{i=1}^{4} v_1(i)}\begin{pmatrix} \frac{d_p r_m r_p + d_m(d_p r_m + s_p(r_m+u_p))}{r_m r_p u_m + (d_m s_m + r_p(s_m+u_m))u_p} \\ \frac{d_p r_m u_m + s_p(r_m u_m + (s_m+u_m)u_p)}{r_m r_p u_m + (d_m s_m + r_p(s_m+u_m))u_p} \\ \frac{d_m s_m(d_p+s_p) + d_p r_p(s_m+u_m)}{r_m r_p u_m + (d_m s_m + r_p(s_m+u_m))u_p} \\ 1 \end{pmatrix}. \tag{2.19}$$

Having performed all the substitutions, the current J becomes rather involved and can be expressed as

$$J = \frac{(u_m r_m d_p r_m) - (d_m s_m u_p s_p)}{J_a + J_b + J_c}, \tag{2.20}$$

with

$$\begin{aligned} J_a &= d_p(r_m(r_p+u_m) + r_p(s_m+u_m)), \\ J_b &= d_m(d_p(r_m+s_m) + r_m s_p + s_m s_p + s_m u_p + s_p u_p), \\ J_c &= (r_p+s_p)(r_m u_m + (s_m+u_m)u_p). \end{aligned}$$

We can finally rewrite the entropy production rate as

$$\sigma = J\log\frac{u_m r_m d_p r_m}{d_m s_m u_p s_p}. \tag{2.21}$$

In order to intuitively understand the expression for σ, we link it to the non-equilibrium properties of our system. In steady state, the general master equation 2.1 satisfies

$$P_i^\infty w_{ij} - P_j^\infty w_{ji} = \pm J, \tag{2.22}$$

with the + (−) sign that holds for all pairs of states where i follows j in the clockwise direction in Fig. 2.1, and J is the steady state current defined in Eq.2.20. The detailed balance condition 2.5 is a special case of Eq. 2.22 where $J = 0$. In order to maintain a non-equilibrium steady state ($J \neq 0$) the system has to dissipate energy at rate $k_B T\sigma$.

We are interested in solving the problem of finding the best network design that can perform a maximally informative delayed readout given a limited and fixed amount of $k_B T\sigma$ units of energy per unit time. This question can be addressed quantitatively by introducing a Lagrange multiplier l, that constrains the energy cost of the transmitted information, and maximizing the functional

$$\mathcal{I}(\tau) - l\frac{\sigma}{\lambda \log 2} = \mathcal{I}(\tau) - l\hat{\sigma} \tag{2.23}$$

over the circuit's reaction rates, $\mathcal{L}$. We rescale the rate of energy dissipation, σ, by the constant $\lambda \log 2$ and call it $\hat{\sigma}$, in order to express both information and entropy production in bits and to measure time in units of the characteristic timescale $1/\lambda$.

For $l = 0$ the constraint on the dissipated energy does not enter the optimization and one recovers the results found in Chap.1 without imposing energetic constraints ($\hat{\sigma} = \sigma = \infty$). In this limit the system is driven out of equilibrium and at least one of the rates vanishes. At the other extreme, when $l = \infty$, any deviation from equilibrium is severely punished and we expect to find the system in equilibrium.

Some intuition about the optimal solutions can be gained before embarking on detailed calculations. In general we can write the probability distribution $P(x_t, z_0)$ as

$$P(x_t, z_0) = \frac{a + bx_t + cz_0 + \mu x_t z_0}{4}. \tag{2.24}$$

The symmetry between the on and off states, $P(+,+) = P(-,-)$ and $P(+,-) = P(-,+)$, implies $b = c = 0$ and normalization $\sum_{\mathrm{i,j}} P(i,j) = 1$ gives $a = 1$. Therefore $P(x_t, z_0)$ has to be of the form

$$P(x_t, z_0) = \frac{1 + \mu x_t z_0}{4}, \tag{2.25}$$

as derived in section 1.4.2.1 [1].

With this form for $P(x_t, z_0)$, we know from section 1.4.2.1 that the mutual information

[1] Another argument for getting Eq.2.25 is to remind the reader that $I[x_t, z_0] < \min\{S[P(x_t)], S[P(z_0)]\}$. So, in order to maximize mutual information, we require the entropy of the input distribution, $S[P(z_0)]$, and of the output distribution, $S[P(x_t)]$, to be both equal to 1 bit. Indeed Eq. 2.25 means $P(x_t) = P(z_0) = 1/2$, which independently maximizes both entropies $S[P(x_t)], S[P(z_0)]$.

in Eq. 2.15 becomes:

$$I = \frac{1+\mu}{2}\log(1+\mu) + \frac{1-\mu}{2}\log(1-\mu), \tag{2.26}$$

where in general $|\mu| \leq 1$. The symmetry of the system results in a degeneracy of solutions, which we break by setting the output flipping rates r_p and r_m to 1. With this choice, the allowed range of μ is $[0, 1]$, and information is a monotonically increasing function of the "effective magnetization" μ and is maximized for $\mu = 1$ giving $I = 1$ bit. We compute μ explicitly for specific models in the following sections.

2.3.2 Simplest model

We start by considering the simplest case depicted in Fig.2.2, where the rates for input z are

$$u_p = u_m = d_p = d_m \equiv u \tag{2.27}$$

and for output x are

$$r_p = r_m \equiv r = 1, \quad s_p = s_m \equiv s. \tag{2.28}$$

Hence we set all the rates for flipping of the input z to be equal, but allow the rates for the output x to be different if the output is aligning with the input ($r_p = r_m \equiv r$) or it is anti-aligning ($s_p = s_m \equiv s$). This models allows z to activate (or repress) x with rate $r(s)$, respectively, but does not allow for feedback since the flipping rate of the input does not depend on the state of the output.

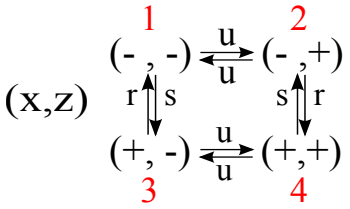


Figure 2.2: The four network states, with corresponding transition rates, in the simplest case where input z can either up or down-regulate the output x but there is no feedback. The input z switches with the same rate regardless of the state of the output x.

The transition rate matrix $\mathcal{L}$ has the form

$$\mathcal{L} = \begin{pmatrix} u+s & -u & -r & 0 \\ -u & u+r & 0 & -s \\ -s & 0 & u+r & -u \\ 0 & -r & -u & u+s \end{pmatrix}. \tag{2.29}$$

Its eigenvalues λ_α are

$$\begin{cases} \lambda_1 = 0 \\ \lambda_2 = 2u \\ \lambda_3 = r+s \\ \lambda_4 = 2u+r+s \end{cases}, \tag{2.30}$$

its right eigevectors v_α are

$$P_\infty = v_1 = \tfrac{1}{2(r+s+2u)} \begin{pmatrix} u+r \\ u+s \\ u+s \\ u+r \end{pmatrix}, \quad v_2 = \frac{1}{2(r+s-2u)} \begin{pmatrix} u-s \\ u-s \\ s-u \\ s-u \end{pmatrix},$$

$$v_3 = \tfrac{1}{2(r+s-2u)} \begin{pmatrix} u-r \\ s-u \\ s-u \\ r-u \end{pmatrix}, \quad v_4 = \frac{u+s}{2(r+s+2u)} \begin{pmatrix} +1 \\ -1 \\ -1 \\ +1 \end{pmatrix}.$$

and its left eigenvectors u_α^T are

$$\begin{cases} u_1^T = (1,1,1,1), \\ u_2^T = (-1, \frac{-u+r}{+u-s}, \frac{+u-r}{+u-s}, 1), \\ u_3^T = (-1,1,-1,1), \\ u_4^T = (1, \frac{-u-r}{+u+s}, \frac{-u-r}{+u+s}, 1). \end{cases}$$

It is possible to write the mutual information in the form of Eq.2.26, where μ is given by

$$\mu = (1-s)\frac{(1+s+2u)e^{-\frac{2u}{\lambda}\tau} - 4ue^{-\frac{(1+s)}{\lambda}\tau}}{(1+s)^2 - 4u^2} \tag{2.31}$$

and time is rescaled with the smallest nonzero eigenvalue λ, as specified in Eq.2.14. The rescaled entropy production rate $\hat{\sigma}$, defined in Eq.2.23, is simply given by

$$\hat{\sigma} = \frac{(1-s)u\log\frac{1}{s}}{\lambda(1+s+2u)}. \tag{2.32}$$

Given that the smallest nonzero can be either $\lambda = 2u$ or $\lambda = r+s = 1+s$, we can define the quantity $\gamma = \frac{1+s}{2u}$ and distinguish two regimes:

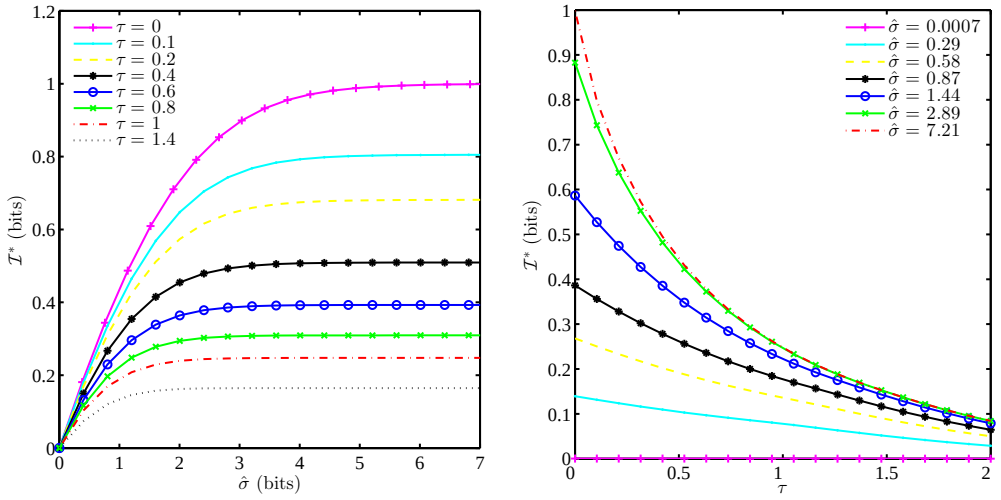
1. $2u < (1+s) \to \gamma > 1$, the output changes on a faster timescale than the input
2. $2u \geq (1+s) \to \gamma \leq 1$, the input changes more quickly than the output.

In general, for each set of rates, the two eigenvalues must be compared and the value of λ (and thus γ) determined.

2.3.2.1 Numerical results

To get an idea about the behavior of the system we will first solve the optimization problem numerically and then interpret the results in terms of some limiting cases. For each readout delay τ and entropy production rate $\hat{\sigma}$, we look for rates that maximize $\mathcal{I}(\tau)$ (given by Eqs. 2.26 and 2.31) and satisfy the expression for $\hat{\sigma}$ (given by Eq. 2.32).

The maximal mutual information values (capacities) of the optimal networks display an intuitive behavior as functions of the dissipated energy and time delay of the readout.



(a) **(b)**

Figure 2.3: Simplest model. (a) Optimal mutual information $\mathcal{I}^*$ as function of entropy production $\hat{\sigma}$, for different delays τ. One can see that information saturates very fast to its asymptotic "ideal" value $\mathcal{I}^*_\infty$, reached for infinite dissipation. Clearly, larger delays entail smaller asymptotic values. (b) Optimal mutual information $\mathcal{I}^*$ as function of the delay τ, for different values of entropy production $\hat{\sigma}$ (measured in bits). The maximum value of information is reached when there is no delay ($\tau = 0$). Such maximum value at $\tau = 0$ decreases linearly with $\hat{\sigma}$.

The mutual information between the input and output of the optimal network decreases with the delay τ of the input readout for all values of dissipation (see Fig. 2.3b), as the network decorrelates. The value of $\mathcal{I}^*$ for $\tau = 0$ decreases linearly with dissipation (see section 2.3.2.2 for derivation). Naturally, the capacities for systems that can dissipate a lot of energy are much larger than those with large energy constraints. However at small time delays the rate of decay of the capacity with time delay is larger for circuits

that function far out of equilibrium than those that are close to equilibrium (see section 2.3.2.3).

Allowing the system to dissipate more energy increases its capacity to transmit information, as it is clear from Fig. 2.3a. Above a certain value of dissipated energy the capacity plateaus and reaches the same value we observed in Chap.1 where we did not constrain dissipation. The value of the this plateau decreases with an increase of the time delay τ of the readout (see section 2.3.2.5 for a functional dependence). The transmitted information decreases to zero linearly with dissipation for all readout delays, $\mathcal{I}^* \sim \frac{c(\tau)^2 \hat{\sigma}}{2\log 2}$, where $c(\tau)$ is a τ dependent constant derived in section 2.3.2.4.

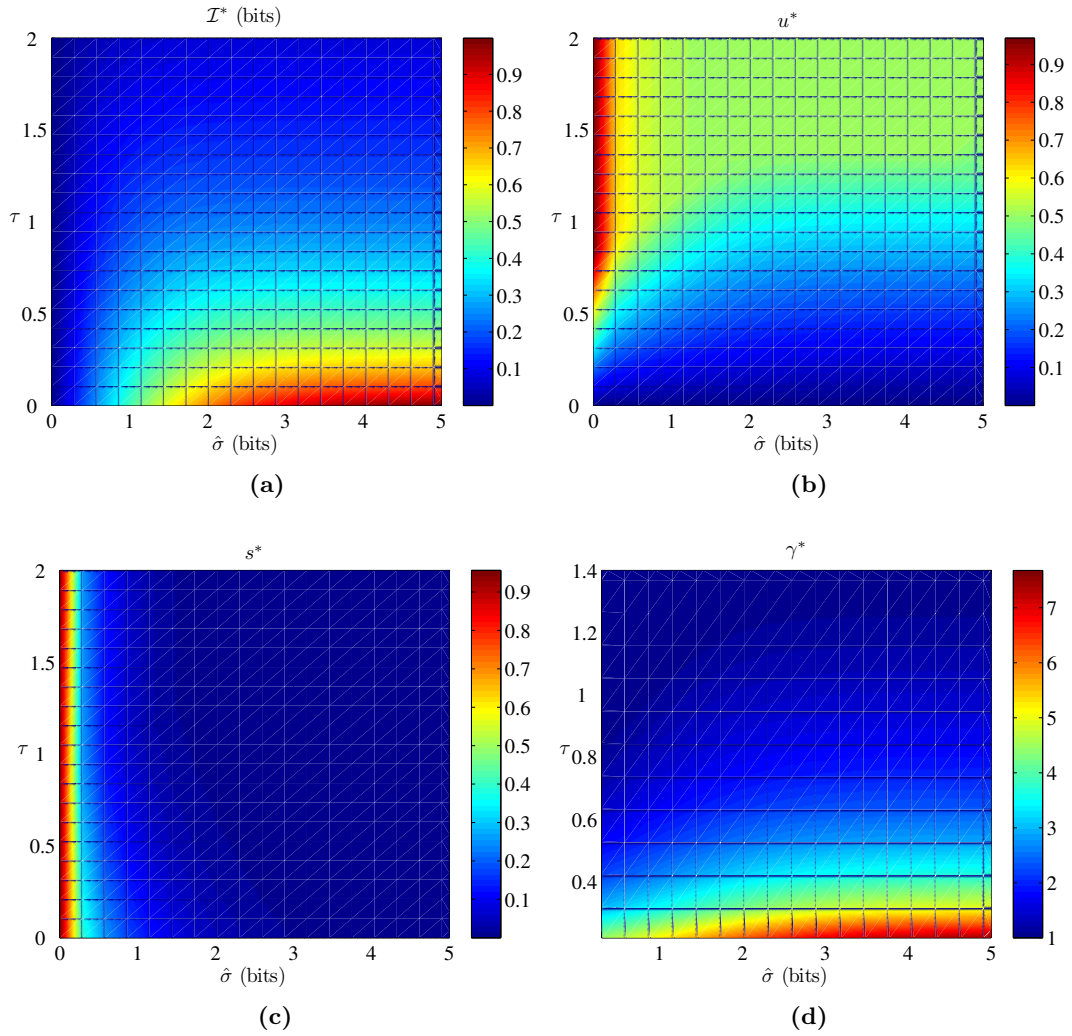


Figure 2.4: Simplest model. (a) Contour plot of optimal mutual information $\mathcal{I}^*$ as function of the readout delay τ and entropy production rate $\hat{\sigma}$. (b) Contour plot of optimal input rate u^* as function of the readout delay τ and entropy production rate $\hat{\sigma}$. (c) Contour plot of optimal output rate s^* as function of the readout delay τ and entropy production rate $\hat{\sigma}$. (d) Contour plot of optimal parameter $\gamma^* = \frac{1+s^*}{2u^*}$ as function of the readout delay τ and entropy production rate $\hat{\sigma}$.

The full behavior of the optimal $\mathcal{I}^*$ in the $(\hat{\sigma}, \tau)$ plane is shown in Fig.2.4a. The different functional dependence on τ and on $\hat{\sigma}$ is very clear from the figure: in particular we stress that, given a fixed delay τ, one cannot transmit information above a certain threshold by simply dissipating more energy. On the other hand, below a certain value of entropy production, the output delay τ barely affects the amount of transmitted information.

In Figures 2.4b and 2.4c we plot the values of the rate constants of the optimal networks that result in the capacities plotted in Fig. 2.4a. We can see that for large dissipation we recover the results of Chap.1: the input rate $u^* = 0$ for $\tau = 0$ and plateaus at 0.5 for any delay $\tau \gtrsim 1.4$, while the output rate $s^* \approx 0$ for any τ. At the other limit where $\hat{\sigma} \ll 1$, rate s^* rapidly goes to 1 when $\hat{\sigma} \to 0$, while u^* is again 0 for $\tau = 0$ but now plateaus at 1 for any delay $\tau \gtrsim 0.8$ (see more detailed discussion in this and in the subsequent sections).

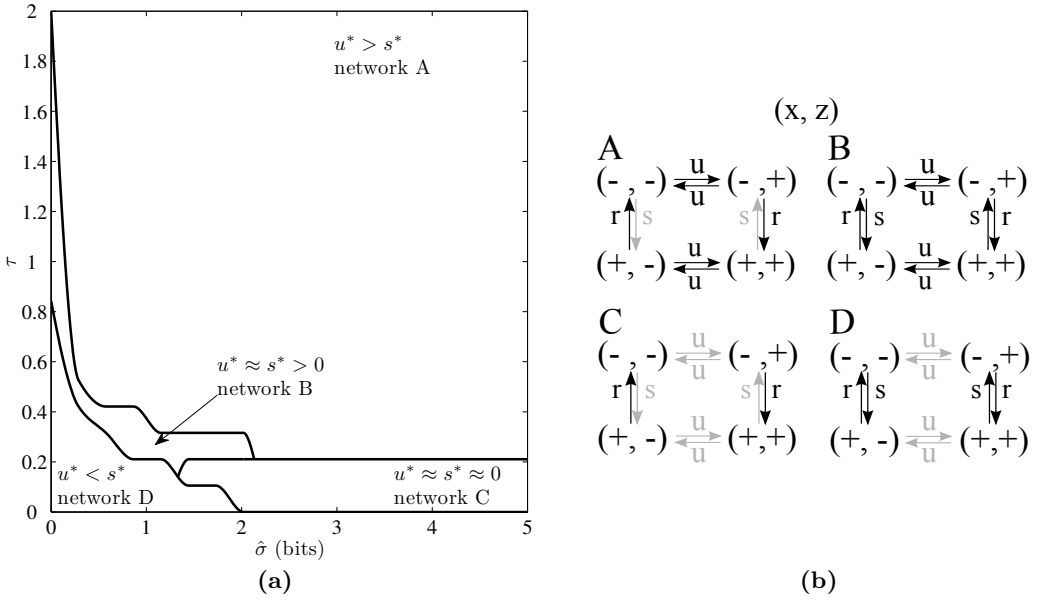


Figure 2.5: Simplest model. (a) Phase diagram in the $(\hat{\sigma}, \tau)$ plane of the optimal network topologies A, B, C, D sketched in panel (b) (gray arrows stand for negligible rates).

To gain a better idea about the optimal network topologies, we have used the rates to broadly classify the circuit topologies in the phase diagram in Fig. 2.5a, with the topologies defined in Fig. 2.5b. In the limit of large dissipation we recover the results we obtained in Chap.1: in the optimal circuit at large readout delays the flipping of the output is governed by an irreversible fast reaction with rate r fixed to 1 (the back reaction is forbidden, that is $s^* = 0$). The output follows the state of the input and the change in the input is described by a reversible slower reaction with rate $u^* < r^*$ (network A in Fig. 2.5b). For shorter delays the flipping rate of the input decreases, causing the capacity to increase. As $\tau \to 0$, $u^* \to 0$ and we obtain two separate subnetworks with

a fixed input in which the output changes quickly to follow the input (network C in Fig. 2.5b).

At large readout delays, the equilibrium solution at $\hat{\sigma} \to 0$ is very similar to the non-equilibrium one, but now detailed balance must always be satisfied. The detailed balanced condition imposes that the output change is completely reversible and thus $s^* \neq 0$. At $\hat{\sigma} = 0$ the forward and back reactions are completely balanced with $s^* = r$ (network B in Fig. 2.5b). Additionally, the input changes on a very fast timescale $u^* \approx r = 1$, faster than for large $\hat{\sigma}$. Not surprisingly, this essentially random flipping in the equilibrium circuit at large delays is not able to reliably transmit information, and $\mathcal{I}^* \approx 0$. For short time delays and small dissipation, $u^* \to 0$ and we obtain two sub-circuits with the output flipping back and forth at the same rate, $s^* = r$ (network D in Fig. 2.5b). Allowing for larger amounts of dissipation breaks detailed balance and decreases the rate of the output's back reaction ($s^* < 1$), so that the output is more likely to be in the same state as the input.

In summary, network C that has a fixed input, which is followed by the output on fast timescales, is the most informative solution. The capacity of this system is reached at finite values of $\hat{\sigma}$, and does not increase further as $\hat{\sigma} \to \infty$. This topology is optimal for a wide range of $\hat{\sigma}$, with the back reaction rate s continuously increasing as the constraints on dissipation impose solutions closer to equilibrium, until network D with the randomly flipping output is reached. At small time delays the optimal solution always keeps the input fixed and adjusts the state of the output to the input ($2u \leq r$). But for large τ the input will change ($2u \sim r$) and the amount of energy that can be dissipated controls whether the output simply follows the input (network A in Fig. 2.5b), or is forced to switch independently (network B in Fig. 2.5b). Information can therefore be lost both in circuits where the output does not have the energy to follow the input (network D) and in circuits where the input decorrelates with time (network A), or where both of these scenarios apply (network B).

Lastly, one can interpret the optimal circuits in terms of the relaxation rate of the system (smallest nonzero eigenvalue). The ratio of the two potentially smallest eigenvalues γ is given by $(1 + s)/(2u)$ – the ratio of the output and the input switching rates. Fig. 2.6 shows the optimal value of γ^*, as a function of the delay τ, in the limit of small entropy production ($\hat{\sigma} = 0.0007$ bits) and of large entropy production ($\hat{\sigma} = 7$ bits). As noted before, for small time delays optimal circuits are those where the input changes more slowly than the output ($\gamma^* > 1$), for all values of dissipation. However for large τ, we define a certain value τ_c at which the input and the output timescales match in optimal circuits, with $\gamma^* = 1$. The value of τ_c depends on the rate of dissipation and corresponds to the optimal rate of input flipping u^* reaching a constant value. For

$\hat{\sigma} \ll 1$, $\tau_c = (1+\sqrt{5})/4$ and $u^* \sim 1$ (see section 2.3.2.4 for derivation). For large dissipation rates $\hat{\sigma} \gg 1$, this delay increases, $\tau_c = (1+\sqrt{3})/2$, and the input switching rate decreases to $u^* = 0.5$ (see section 2.3.2.5 for derivation). We already know from Chap.1 that the system's optimality is attained, for small delays τ, when the output is *faster* than the input, while for delays $\tau \geq \tau_c$ when the input and the ouput timescales *match*. This matching of timescales is possible at $\tau \geq \tau_c > 0$ even if the system cannot dissipate energy (small $\hat{\sigma}$). On the other hand, even when the system has no energetic constraints ($\sigma \to \infty$), it picks nevertheless a finite $\tau_c > 0$.

Finally, the optimal solution is always in the $\gamma^* \geq 1$ limit, where the input changes more slowly than the output, as it is clear from the contourplot 2.4d.

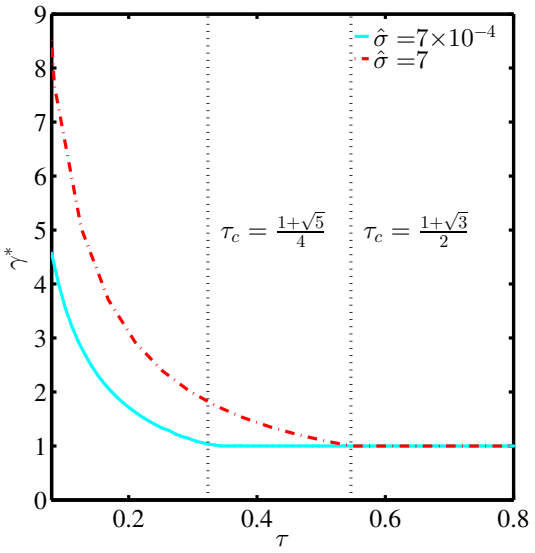


Figure 2.6: The quantity $\gamma = (1+s)/(2u)$, is the ratio of two smallest nonzero eigenvalues, corresponding to the output and the input timescale, respectively. The optimal value γ^* is shown as a function of delay τ, in the two limits of very small and very large entropy production $\hat{\sigma}$ (measured in bits). The time τ_c after which the output timescale matches the input timescale ($\gamma^* = 1$) decreases with dissipation from $1+\sqrt{3}/2$ when $\hat{\sigma} \to \infty$ to $1+\sqrt{5}/4$ when $\hat{\sigma} \to 0$.

Having understood the general behavior of the capacity for this model, we can exploit its simplicity to obtain precise analytical scaling results in the limits of small and large delay and dissipation.

2.3.2.2 Limit $\tau = 0$

The simplest case is that of instantaneous readout, $\tau = 0$, where the effective magnetization μ is

$$\mu = \left(\frac{1-s}{1+s}\right)\frac{\gamma}{\gamma+1}. \tag{2.33}$$

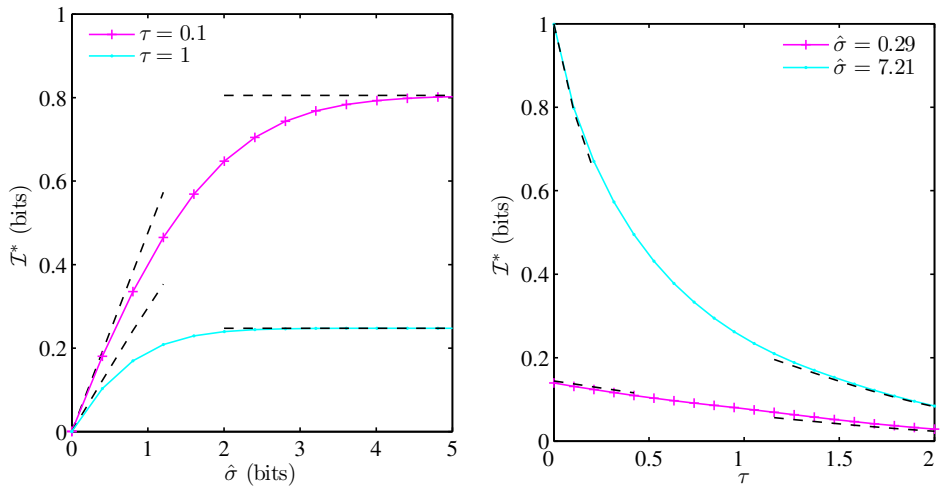


Figure 2.7: Comparison of the analytical (dashed lines) and numerical solutions (solid lines) for optimal mutual information $\mathcal{I}^*$. In panel (a) the dependence on entropy production $\hat{\sigma}$ is shown for $\tau = 0.1$ and for $\tau = 1$. In panel (b) the dependence on τ is represented for $\hat{\sigma} = 0.29$ bits and $\hat{\sigma} = 7.21$ bits. The analytical expansions are the ones derived in sections 2.3.2.3, 2.3.2.4 and 2.3.2.5.

We can formally rewrite Eq. 2.32 as:

$$s = \exp\left[-2\beta(\hat{\sigma}, \gamma)\hat{\sigma}\right], \tag{2.34}$$

where $\beta(\hat{\sigma}, \gamma)$ is in general a nonlinear function of $\hat{\sigma}$ and γ. This form agrees with the numerical results for s^* that show a strong decay with $\hat{\sigma}$ (Fig. 2.4c). Moreover, with this parametrization when $\hat{\sigma} \to \infty$ we recover the optimal solution $s^* = 0$ of Chap.1, whereas for $\hat{\sigma} = 0$ one correctly gets $s = 1$ from Eq.2.32 [2].

As we know from our numerical exploration, in the $\tau = 0$ limit the capacity strongly depends on the value of $\hat{\sigma}$. First we can explore the limit of large dissipation, where we know from Chap.1 (and from the results presented in Fig. 2.4c) that s^* is small. In this limit Eq. 2.34 simplifies (Eq. 2.32 is explicitly solved for s) and β is a function of only u and λ, not of $\hat{\sigma}$.

To find γ^* that maximizes μ, we exploit the parametrization 2.34 of s to write

$$\mu = \tanh\left(\beta(u, \lambda)\hat{\sigma}\right) \frac{\gamma}{1+\gamma}. \tag{2.35}$$

[2] In the regime where $\lambda = 2u$ the implication $\hat{\sigma} = 0 \Leftrightarrow s = 1$ is straightforward. On the other hand, if $\lambda = 1 + s$, in principle a null value of $\hat{\sigma}$ could also be obtained setting $u = 0$, but this would mean that $2u < 1 + s$ and thus $\lambda = 2u$. So the only consistent solution demands $s = 1$.

At fixed $\hat{\sigma}$, the largest value of μ is always achieved for $\gamma^* = \infty$. This means the output changes on faster timescales than the input and the smallest eigenvalue is $\lambda = 2u$. More precisely, as in the dissipation-less case of Chap.1, the optimal rate is $u^* = 0$ at $\tau = 0$.

To find s^*, we substitute the parametrization 2.34 for s into Eq. 2.32 with $\lambda = 2u$ and obtain at fixed $\hat{\sigma}$:

$$1 + \frac{1}{\gamma} = \beta \tanh(\beta\hat{\sigma}). \tag{2.36}$$

Since $\gamma^* = \infty$, β^* must satisfy

$$\beta^* \tanh(\beta^*\hat{\sigma}) = 1. \tag{2.37}$$

Hence, for large dissipation rates, $\beta^* \sim 1$, $s^* \sim e^{-2\hat{\sigma}}$ is exponentially small and $\mu^* \sim \tanh(\hat{\sigma}) \sim 1$. This results in the optimal information $\mathcal{I}^* \sim 1$ bit.

For small dissipation rates, Eq. 2.37 becomes $\beta^*(\hat{\sigma}) \sim 1/\sqrt{\hat{\sigma}}$, and Eq. 2.35 results in the effective magnetization $\mu \sim \beta^*(\hat{\sigma})\hat{\sigma} \sim \sqrt{\hat{\sigma}}$. Finally, the optimal mutual information goes to 0 linearly with the rescaled dissipation, $\mathcal{I}^* \approx (\mu^*)^2/2 \simeq \hat{\sigma}/2$ bits.

2.3.2.3 Limit $\tau \ll 1$

The results from the $\tau = 0$ limit serve as a basis for considering the scaling of the mutual information for small but finite $\tau \ll 1$. Since γ^* diverges at $\tau = 0$, we assume that for $\tau \to 0$ the smallest eigenvalue is still $2u$ and $\gamma^* > 1$. We will also use the generalized nonlinear parametrization of s in Eq. 2.34, as we did for $\tau = 0$.

In the small dissipation limit $\hat{\sigma} \ll 1$, Eq. 2.36 becomes:

$$\beta \sim \frac{1}{\sqrt{\hat{\sigma}}}\sqrt{\frac{\gamma+1}{\gamma}}. \tag{2.38}$$

Using Equations 2.34,2.38 and the fact that $\lambda = 2u$, the effective magnetization of Eq. 2.31 becomes a function of γ and $\hat{\sigma}$ only:

$$\mu = \sqrt{\hat{\sigma}}\sqrt{\frac{\gamma+1}{\gamma}}\frac{\gamma}{\gamma^2-1}[(\gamma+1)e^{-\tau} - 2e^{-\gamma\tau}]. \tag{2.39}$$

We maximize the effective magnetization with respect to γ, $d\mu/d\gamma = 0$, and assume the scaling

$$\gamma^* \simeq \frac{a_0}{\tau} + b_0 + c_0\tau. \tag{2.40}$$

Solving the resulting equations for the coefficients in orders of τ (see Appendix B.1.1 for details), the optimal effective magnetization is

$$\mu^* \sim \sqrt{\hat{\sigma}}(1 + A_0\tau), \tag{2.41}$$

where $A_0 = -0.24...$ is computed exactly in Appendix B.1.1.

In the large dissipation limit $\hat{\sigma} \gg 1$, Eq. 2.36 becomes

$$\beta \simeq \frac{\gamma + 1}{\gamma}. \tag{2.42}$$

Using Eq. 2.34, $\lambda = 2u$ and the fact that in this limit $s \to 0$, the effective magnetization in Eq. 2.31 is

$$\mu = \frac{\gamma}{\gamma^2 - 1}[(\gamma + 1)e^{-\tau} - 2e^{-\gamma\tau}]. \tag{2.43}$$

Assuming $\gamma^* \simeq \frac{a_\infty}{\tau} + b_\infty + c_\infty\tau$, a calculation analogous to the small dissipation limit results in the maximized effective magnetization

$$\mu^* \sim 1 + A_\infty\tau + B_\infty\tau^2, \tag{2.44}$$

where $A_\infty = -0.63...$ and $B_\infty = 0.23...$ are computed exactly in Appendix B.1.1.

Summarizing, in the small dissipation limit we find

$$\mathcal{I}^* \approx \frac{\mu^{*2}}{2} \sim \frac{\hat{\sigma}(1 + 2A_0\tau)}{2}, \tag{2.45}$$

a linear scaling of the information both with dissipation and with readout delay. In the large dissipation limit the information is independent of the dissipation and $\mathcal{I}^* \to 1$ bit, as μ^* tends to one quadratically in the delay, as given by Eq. 2.44.

All these scaling behaviors are shown as broken lines and compared with numerical results (solid lines) in Fig. 2.7.
In particular, Eq.2.45 is plotted in panel 2.7a in the range of $\hat{\sigma} \ll 1$ and for $\tau = 0.1$, and in panel 2.7b in the range of $\tau \ll 1$ and for $\hat{\sigma} = 0.29$ bits.
On the other hand, the optimal $\mathcal{I}^*(\mu^*)$, with μ^* given in Eq.2.44, is shown in panel 2.7a in the range of $\hat{\sigma} \gg 1$ for $\tau = 0.1$.

2.3.2.4 Limit $\hat{\sigma} \ll 1$

Having explored the behavior of the system for small delays τ, we can now investigate the limit of $\hat{\sigma} \to 0$ at any τ: in order to do so, we recall from Eq.2.34 that in this limit rate s goes to 1. Hence we can set $s = 1 - \epsilon$ and expand in ϵ to leading order.

We start from considering the regime in which $\lambda = 2u$ and hence $\gamma \geq 1$: using Eq.2.32 we get for the entropy production

$$\hat{\sigma} \simeq \frac{\epsilon^2 \gamma}{4(1+\gamma)\log 2} \quad i.e. \quad \epsilon \simeq 2\sqrt{\hat{\sigma}\frac{1+\gamma}{\gamma}\log 2}, \tag{2.46}$$

while the magnetization of Eq. 2.31 reads

$$\mu \simeq \frac{\epsilon}{2}\gamma \frac{(1+\gamma)e^{-\tau} - 2e^{-\gamma\tau}}{\gamma^2 - 1}. \tag{2.47}$$

Substituting ϵ into the expression for μ reads $\mu \simeq c_+(\gamma,\tau)\sqrt{\hat{\sigma}}$, where:

$$c_+(\gamma,\tau) = \sqrt{\frac{\gamma \log 2}{\gamma+1}} \frac{-2e^{-\gamma\tau} + (1+\gamma)e^{-\tau}}{\gamma - 1}. \tag{2.48}$$

We then consider the regime with $\lambda = 1 + s$ and $\gamma < 1$: we now get

$$\hat{\sigma} \simeq \frac{\epsilon^2}{4(1+\gamma)\log 2} \quad i.e. \quad \epsilon \simeq 2\sqrt{\hat{\sigma}(\gamma+1)\log 2}, \tag{2.49}$$

and

$$\mu \simeq \frac{\epsilon}{2}\gamma \frac{-2e^{-\tau} + (1+\gamma)e^{-\tau/\gamma}}{\gamma^2 - 1}. \tag{2.50}$$

Again, by substituting ϵ into Eq.2.50, we find $\mu \simeq c_-(\gamma,\tau)\sqrt{\hat{\sigma}}$ with

$$c_-(\gamma,\tau) = \sqrt{\frac{\log 2}{\gamma+1}}\gamma \frac{-2e^{-\tau} + (1+\gamma)e^{-\tau/\gamma}}{\gamma - 1}. \tag{2.51}$$

In order to maximize the mutual information $\mathcal{I} \propto \mu^2$, we need to solve for

$$c(\tau) = \max\{\max_{\gamma \geq 1} c_+(\gamma,\tau), \max_{\gamma < 1} c_-(\gamma,\tau)\}. \tag{2.52}$$

The coefficient $c_-(\gamma,\tau)$ is an increasing function of $\gamma \in [0,1)$ for any τ. Hence we need to consider the regime where $\gamma \in [1,\infty)$ and maximize $c_+(\gamma,\tau)$. For each value of τ, such function has a single maximum in γ, which is a decreasing function of τ and satisfies the transcendental equation

$$e^{(\gamma^*-1)\tau} = 2\frac{1+\gamma^*+2\gamma^{*2}+2\gamma^*\tau(\gamma^{*2}-1)}{(1+\gamma^*)(1+3\gamma^*)}. \tag{2.53}$$

In the limit $\gamma^* \to 1^+$, the delay τ_c which maximizes $c_+(\gamma^*,\tau)$ is found as the solution of

$$e^{-\tau}\frac{(1+2\tau-4\tau^2)}{4\sqrt{2}} = 0, \tag{2.54}$$

and reads

$$\tau_c = \frac{1+\sqrt{5}}{4}. \tag{2.55}$$

For all larger values of τ, the optimal γ^* sticks at 1 and the proportionality coefficient $c(\tau) = c_+(1,\tau)$ is simply

$$c(\tau) = \frac{e^{-\tau}(1+2\tau)}{\sqrt{2}}. \tag{2.56}$$

The optimal mutual information $\mathcal{I}^*$ is linear in dissipation and exponentially decaying in τ:

$$\mathcal{I}^* \simeq \frac{c(\tau)^2\hat{\sigma}}{2\log 2}. \tag{2.57}$$

It is shown as broken lines and compared with numerical results (solid lines) in Fig. 2.7. In particular, in panel 2.7a it is plotted in the range of $\hat{\sigma} \ll 1$ for $\tau = 1$, while in panel 2.7b it is displayed in the range of large τ for $\hat{\sigma} = 7.21$ bits.

2.3.2.5 Limit $\hat{\sigma} \gg 1$

In this regime, we can extend the observations from the small τ limit to postulate that the effective magnetization is weakly dependent on the entropy production:

$$\mu \simeq c_+(\gamma,\tau) = \frac{\gamma}{\gamma^2-1}[-2e^{-\gamma\tau} + (\gamma+1)e^{-\tau}], \quad \gamma \geq 1 \tag{2.58}$$

$$\mu \simeq c_-(\gamma,\tau) = \frac{\gamma}{\gamma^2-1}[-2e^{-\tau} + (\gamma+1)e^{-\tau/\gamma}], \quad \gamma < 1 \tag{2.59}$$

The expression 2.59 is maximized always at the boundary $\gamma^* = 1$, while the expression 2.58 has a single maximum in γ for each τ. Such γ is a decreasing function of τ and satisfies the trascendental equation

$$e^{(\gamma-1)\tau} = \frac{2(1+\gamma^2+\gamma\tau(\gamma^2-1)}{(1+\gamma)^2}. \tag{2.60}$$

Similar considerations as in the small dissipation case result in

$$\tau_c = \frac{1+\sqrt{3}}{2}, \tag{2.61}$$

above which the optimal γ^* sticks at 1 and the magnetization simply reads

$$\mu \simeq c(\tau) = \frac{e^{-\tau}(1+2\tau)}{2}. \tag{2.62}$$

Finally, the optimal mutual information approaches a plateau for large $\hat{\sigma}$ given by

$$\mathcal{I}^* \simeq \begin{cases} 1 + \tau\tilde{a}\log_2(\tau\tilde{a}/e), & \gamma^* \gg 1, \tau \ll 1 \\ \frac{c(\tau)^2}{2\log 2}, & \gamma^* = 1, \tau \gg 1 \end{cases} \tag{2.63}$$

The coefficient $\tilde{a} = 0.31...$ is defined as $\frac{a_\infty}{2(a_\infty+1)}$, with a_∞ introduced in section 2.3.2.3. The above limiting behavior of $\mathcal{I}^*$ is plotted as broken lines and compared with the numerical result (solid line) in Fig. 2.7b for $\hat{\sigma} = 7.21$ bits.

2.3.3 Adding feedback

We now ask how allowing for *feedback* between the output and the input changes the energetic constraints on the optimally informative solutions. In terms of our model, this corresponds to saying that the input switching rates depend on the state of the output ($u_p \neq u_m$ and $d_p \neq d_m$), unlike in the simplest case of the model discussed in section 2.3.2. In circuits with feedback and without additional inputs the difference between the input and output is no longer clear: z is an input for x and vice versa.

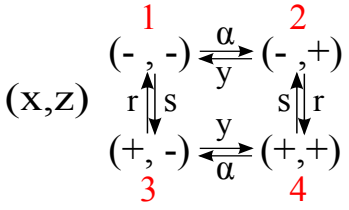


Figure 2.8: The four network states, with corresponding transition rates, in a model with feedback where the input z rates depend on the state of the output variable x.

We can exploit the symmetry between + and − states to decrease the number of rates in the network and set the rates of aligning (and anti aligning) of the input to the output to be equal, regardless of the state of the input. Specifically, the rates defined in Fig. 2.1 simplify to the ones shown in Fig. 2.8, that is

- $r_p = r_m \equiv r = 1$,
- $s_p = s_m \equiv s \leq 1$,
- $d_p = u_m \equiv \alpha \leq 1$,
- $d_m = u_m \equiv y \leq 1$.

We know from Chap.1 that in the infinite dissipation limit the optimal solutions cycle irreversibly through the four states in a clockwise direction.

In terms of the rates defined in Fig. 2.8 the rate matrix $\mathcal{L}$ now reads

$$\mathcal{L} = \begin{pmatrix} s+\alpha & -y & -1 & 0 \\ -\alpha & 1+y & 0 & -s \\ -s & 0 & 1+y & -\alpha \\ 0 & -1 & -y & s+\alpha \end{pmatrix}. \tag{2.64}$$

It is useful to introduce the quantities A and ρ, defined as

$$A = 1+s+y+\alpha \tag{2.65}$$

$$\rho = \sqrt{(1+s+y+\alpha)^2 - 8(sy+\alpha)}. \tag{2.66}$$

Then the eigenvalues λ_α can be written as

$$\begin{cases} \lambda_1 = 0 \\ \lambda_2 = A \\ \lambda_3 = \frac{1}{2}(A-\rho) \\ \lambda_4 = \frac{1}{2}(A+\rho) \end{cases}. \tag{2.67}$$

It is clear that the smallest nonzero eigenvalue is always $\lambda = \frac{A-\rho}{2}$.

The right eigenvectors v_α are given by

$$P_\infty = v_1 = \frac{1}{2A}\begin{pmatrix} 1+y \\ s+\alpha \\ s+\alpha \\ 1+y \end{pmatrix}, \tag{2.68}$$

$$v_2 = \frac{s+\alpha}{2A}\begin{pmatrix} +1 \\ -1 \\ -1 \\ +1 \end{pmatrix}, \tag{2.69}$$

$$v_3 = \frac{1}{4\rho}\begin{pmatrix} +(-1+s-y+a-\rho) \\ +2(s-\alpha) \\ -2(s-\alpha) \\ -(-1+s-y+a-\rho) \end{pmatrix}, \tag{2.70}$$

$$v_4 = \frac{1}{4\rho}\begin{pmatrix} +(1-s+y-\alpha-\rho) \\ +2(-s+\alpha) \\ -2(-s+\alpha) \\ -(1-s+y-\alpha-\rho) \end{pmatrix}, \tag{2.71}$$

and the left eigenvectors u_α are

$$u_1^T = (1,1,1,1), \tag{2.72}$$
$$u_2^T = (1,-\frac{1+y}{s+\alpha},-\frac{1+y}{s+\alpha},1), \tag{2.73}$$
$$u_3^T = \left(-1,\frac{2(1-y)}{1-s+y-\alpha+\rho},-\frac{2(1-y)}{1-s+y-\alpha+\rho},1\right), \tag{2.74}$$
$$u_4^T = \left(-1,\frac{2(1-y)}{1-s+y-\alpha-\rho},-\frac{2(1-y)}{1-s+y-\alpha-\rho},1\right). \tag{2.75}$$

After rescaling by the smallest eigenvalue $\lambda = \frac{A-\rho}{2}$, the entropy production reads

$$\hat{\sigma} = \frac{2(\alpha-sy)}{A(A-\rho)}\log_2\left(\frac{\alpha}{sy}\right). \tag{2.76}$$

The mutual information is expressed by Eq.2.26 in terms of the effective magnetization

$$\mu = \exp\left(-\frac{A}{2\lambda}\tau\right)\left\{q\cosh\left(\frac{\rho}{2\lambda}\tau\right) - \frac{\left[s^2-(1+y)^2-4\alpha+\alpha^2+2s(2y+\alpha)\right]}{A\rho}\sinh\left(\frac{\rho}{2\lambda}\tau\right)\right\}, \tag{2.77}$$

with $q = (1+y-s-\alpha)/A$ and time rescaled by the smallest nonzero eigenvalue λ, as explained in Eq. 2.14 (see Appendix B.2.1 for a detailed calculation of μ).

Despite the analytic expressions, the nonlinearities of the problem prohibit finding analytical solutions to this constrained optimization problem. However, we can explore analytically some limiting cases which provide interesting insights.

2.3.3.1 Limit $\hat{\sigma} = 0$

We start from noticing that, if the rates satisfy the condition

$$\alpha = sy, \tag{2.78}$$

then the entropy production of Eq.2.76 is identically zero, whereas the magnetization μ of Eq.2.77 can be cast in a simple form which allows us to optimize it analytically.

We start from the expression

$$\mu = \frac{q}{2\rho}\left(e^{-\tau}(A+\rho) - e^{-\left(\frac{A+\rho}{A-\rho}\right)\tau}(A-\rho)\right), \tag{2.79}$$

obtained from Eq.2.77 by setting $\alpha = sy$ and doing some algebraic manipulations. Then we reparametrize the rates as

$$w = \frac{4s}{(1+s)^2}, \quad v = \frac{4y}{(1+y)^2} \tag{2.80}$$

and define also

$$\phi = \frac{1+\sqrt{1-wv}}{1-\sqrt{1-wv}}. \tag{2.81}$$

Then ρ is simply

$$\rho = A\sqrt{1-wv} \tag{2.82}$$

and the final expression for μ is

$$\mu = \frac{\sqrt{1-w}}{2\sqrt{1-wv}}\left(e^{-\tau} - e^{-\phi\tau}\right) + \frac{\sqrt{1-w}}{2}\left(e^{-\tau} + e^{-\phi\tau}\right). \tag{2.83}$$

When $\tau = 0$, the magnetization is simply $\mu = \sqrt{1-w} = \frac{1-s}{1+s}$. It is evident that μ is maximized when $w^* = 0$ (that is $s^* = 0$ and hence $\alpha^* = s^* y = 0$) at all times, reading $\mu^* = e^{-\tau}$. The optimal mutual information $\mathcal{I}^*$ thus attains 1 bit for $\tau = 0$ and decays in time as

$$\mathcal{I}^* = \frac{1}{2}\left(\log_2(1-e^{-2\tau}) + e^{-\tau}\log_2\frac{1+e^{-\tau}}{1-e^{-\tau}}\right). \tag{2.84}$$

The two "mixed" states $(x,z) = \{(+,-),(-,+)\}$ are not accessible, while the two "aligned" states $(+,+)$ and $(-,-)$ have probability $1/2$ (see Eq.2.68 for the steady-state probability vector P^∞).

This optimal solution corresponds to a completely unresponsive network, where there are no fluxes. The value of the nonzero rate y does not matter, since it leads from completely forbidden states. The highest possible value of information transmission is guaranteed, while remaining in an equilibrium configuration in which detailed balance is satisfied and there is no regulation. If the readout occurs at later times, the transmitted information decays, however the nature of the solution remains the same. In summary, the optimal solution in equilibrium corresponds to a "dead" system, which is very informative, but not necessary useful.

It is much more intriguing to consider the suboptimal solution where $s = \eta$ and $\alpha = \eta y$, with $\eta \ll 1$: now the "mixed" states have probability $\frac{\eta}{(1+y)(1+\eta)}$ and the full network is recovered. The system is able to cycle through the four states and transmit almost 1 bit of information, without dissipating energy.

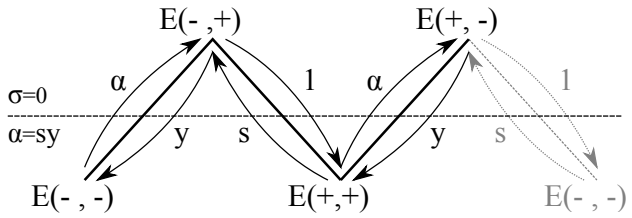


Figure 2.9: Sketch of the network in Fig.2.8, now endowed with an energetic description. Here we show the limiting case where entropy production σ is zero and transition rates are related by the condition $\alpha = sy$, which implies $E(+,-) = E(-,+) > E(+,+) = E(-,-)$.

In order to better understand the energetic framework, we make use of the detailed-balance condition 2.5 and of the Boltzmann relation, $P_i = \exp(-E_i/k_BT)$, between the probability P_i and the energy E_i of state i. Then we can write the following conditions:

$$\alpha/y = \epsilon = e^{-\frac{E(+,-)-E(+,+)}{k_BT}}, \tag{2.85}$$

$$s = \epsilon = e^{-\frac{E(-,+)-E(+,+)}{k_BT}}, \tag{2.86}$$

$$\alpha/y = \epsilon = e^{-\frac{E(-,+)-E(-,-)}{k_BT}}, \tag{2.87}$$

$$s = \epsilon = e^{-\frac{E(+,-)-E(-,-)}{k_BT}}, \tag{2.88}$$

$$\tag{2.89}$$

which immediately imply that

$$E(+,-) = E(-,+) > E(+,+) = E(-,-). \tag{2.90}$$

The above considerations are pictured in Fig.2.9: we can see that, as long as only two finite energetic levels are present - one for the "mixed" states and one for the "aligned" states - the system is able to cycle indefinitely among them at no cost. Input z can up and down-regulate output x, thus transmitting information. At $s = 0$ (which implies $\alpha = 0$ from the equilibrium condition) infinite energy barriers separate the aligned states and lead to the unresponsive dead solution. When $s > 0$, input z controls output x, transmitting information. It is worth stressing that such costless informative solution is present only because of feedback: indeed, in the simpler model of section 2.3.2, rates α and y are equal to u and $s = 1$ is the only solution of the dissipation-less condition $\alpha = sy$, thus forcing $\mathcal{I}$ to be zero.

2.3.3.2 Limit $\hat{\sigma} \ll 1$

After having explored the singular case with $\sigma \equiv 0$, we can ask ourselves what is the system's optimal behavior when the rescaled entropy production $\hat{\sigma}$ is nonzero but very

small. In order to do so, we expand α around the dissipationless solution, as $\alpha = sy(1-\epsilon)$, with $\epsilon \ll 1$. For the sake of tractability, we limit ourselves to the case with $\tau = 0$, where

$$\mu = \frac{1 - s + y - \alpha}{1 + s + y + \alpha}. \tag{2.91}$$

The rescaled dissipation 2.76 in terms of ϵ and of the parametrization 2.80 is

$$\hat{\sigma} \simeq \frac{wv\epsilon^2}{8(1 - \sqrt{wv}) \log 2}, \tag{2.92}$$

where we have kept only the leading order term in ϵ. Solving for $\epsilon > 0$ reads

$$\epsilon = 2\sqrt{2\hat{\sigma}\frac{1 - \sqrt{1 - wv}}{wv} \log 2}. \tag{2.93}$$

When expanded to first order in ϵ, the magnetization μ of Eq.2.91 is simply given by

$$\mu \simeq \frac{1 - s}{1 + s} + \frac{2sy\epsilon}{(1 + s)^2(1 + y)}. \tag{2.94}$$

After substituting the solution 2.93 for ϵ and exploiting the parametrizations 2.80, we finally get the following relationship between μ and $\hat{\sigma}$:

$$\mu \simeq \sqrt{1 - w} + \sqrt{\hat{\sigma} \log 2}\frac{\sqrt{wv(1 - \sqrt{1 - wv})}}{\sqrt{2}(1 + \sqrt{1 - v})}. \tag{2.95}$$

In order to optimize μ at fixed $\hat{\sigma}$, we clearly need to take w as small as possible (hence $s \approx 0$), so that the term $\sqrt{1 - w}$ is close to 1. Then, it is pretty straightforward to show that taking $v = 1$ (hence $y = 1$) maximizes the coefficient of $\sqrt{\hat{\sigma}}$. Combining the two terms, it is possible to get $\mu^* \approx 1$ and consequently $\mathcal{I}^* \approx 1$.

Unlike in the model without feedback, for small delays τ it is possible to achieve almost 1 bit of information even when entropy production is arbitrary small, thanks to the fact that the back and forward switching rates of the input may now differ from each other ($\alpha \neq y$). This allows the system to spend most of its time in the aligned states $(+,+)$ and $(-,-)$, and thus to make the readout of the input by the output as informative as possible.

At the same time, since all rates are larger than 0, the network features all the four states. Such a solution is already present as a suboptimal one when $\hat{\sigma} = 0$, that is when the optimal network is unresponsive and only the two aligned states are visited ($s^* = \alpha^* = 0$). Basically, "switching on" dissipation allows the system to become responsive while being still maximally informative. However, since rates s^* and α^* are close to 0, the flow of probability through the four states, while theoretically present, is very small and the system practically behaves as in the dissipationless optimal case.

2.3.3.3 Limit $\hat{\sigma} \gg 1$

When entropy production is very large, we know from Chap.1 that mutual information is maximized when rates s^* and y^* are equal to zero for all delays τ. Indeed the expression 2.76 for $\hat{\sigma}$ diverges for such values of s and y.

In order to find α^*, we can write the magnetization μ in terms of α and τ, after having set $s = y = 0$:

$$\mu(\alpha,\tau) = \frac{1}{2(1+\alpha)\rho}\left(e^{-\tau\frac{(1+\alpha+\rho)^2}{1-\alpha+\rho}} + e^{-\tau\frac{1+\alpha+\rho}{1+\alpha-\rho}\frac{(1+\alpha-\rho)^2}{-1+\alpha+\rho}}\right), \tag{2.96}$$

where $\rho = \sqrt{1+\alpha^2-6\alpha}$. Expanding μ to the first order in τ

$$\mu_{\tau\ll 1}(\alpha,\tau) \simeq \frac{1-\alpha}{1+\alpha} + \frac{(1+\alpha+\rho)\,\tau}{2(1+\alpha)}, \tag{2.97}$$

we find α^* that maximizes the above expression:

$$\alpha^*(\tau) = \frac{(1-\tau)\tau}{2-\tau}. \tag{2.98}$$

α^* is an increasing function of τ, until it reaches the value $\alpha_c^* = 3-2\sqrt{2}$ when $\tau_c = 2-\sqrt{2}$. For such value of α_c^*, $\rho = 0$ and the two smallest eigenvalues $(A-\rho)/2$ and $(A+\rho)/2$ become degenerate. Values of α larger than α_c^* are not optimal, since then ρ would becomes complex and oscillations would be detrimental for information transmission (see section 1.4.2.4). The values of α_c^* and τ_c are exactly those we found in the previous Chapter (section 1.4.2.4). The optimal mutual information $\mathcal{I}^*$ as a function of τ is obtained as the hull curve of $\mathcal{I}(\alpha^*(\tau),\tau)$ and $\mathcal{I}(\alpha_c^*,\tau)$.

2.3.3.4 Numerical results

Having discussed some analytically tractable limits, we can now move to a full numerical optimization. Our approach is to divide the optimization into two "branches": in one branch (called YA) we constrain $y \geq \alpha$, in the other branch (called AY) we force $y \leq \alpha$. In this way we explore two different topologies: in the first one the system cycles in the CCW direction in Fig.2.8, while in the second one the system cycles through the four states in the CW direction, recalling the findings of section 1.4.2.4 for infinite energy dissipation.

Optimal mutual information $\mathcal{I}^*$ is shown in Figures 2.10a and 2.10b as function of entropy production $\hat{\sigma}$ and delay τ, respectively.

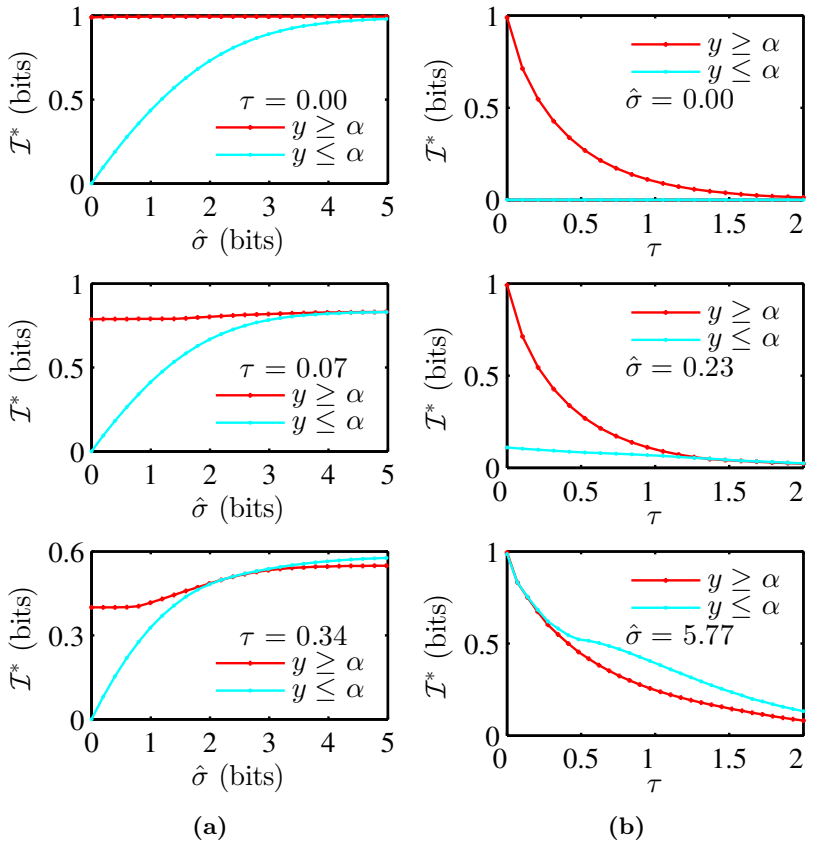


Figure 2.10: Adding feedback. 2.10a Optimal mutual information $\mathcal{I}^*$ as function of rescaled dissipation $\hat{\sigma}$. Values for $\tau = 0$, $\tau = 0.07$ and $\tau = 0.34$ are displayed at the top, center and bottom, respectively. Results from the simulation branch with $y \geq \alpha$ and with $y \leq \alpha$ are plotted as dotted red lines and solid cyan lines, respectively. Rates used to compute such mutual information are shown in Fig.B.1. 2.10b Optimal mutual information $\mathcal{I}^*$ as function of delay τ. Values for $\hat{\sigma} = 0$ bits, $\hat{\sigma} = 0.23$ bits and $\hat{\sigma} = 5.77$ bits are displayed at the top, center and bottom, respectively. Results from the simulation branch with $y \geq \alpha$ and with $y \leq \alpha$ are plotted as dotted red lines and solid cyan lines, respectively. Rates used to compute such mutual information are shown in Fig.B.2. The solutions of the two branches $y \geq \alpha$ and $y \leq \alpha$ coincide at large $\hat{\sigma}$ and small τ and at small $\hat{\sigma}$ and large τ. This happens because the back and forward input flipping rates are equal: in the first case $y^* \approx \alpha^* \approx 0$, while in the second case $y^* \approx \alpha^* > 0$ (see Fig. 2.12a).

Let us begin from Fig.2.10a: here we report the dependence of $\mathcal{I}^*$ on the rescaled dissipation $\hat{\sigma}$, for different delays τ. Results of the optimization branches YA and AY are plotted as dotted red lines and solid cyan lines, respectively. Optimal rates used to calculate $\mathcal{I}^*$ are shown in Fig.B.1 in Appendix B.

For $\tau = 0$ we see that the branch YA delivers ≈ 1 bit of information even for very small $\hat{\sigma}$, as shown analytically in section 2.3.3.2. The branch AY is instead suboptimal for any finite entropy production rate and matches branch YA only when $\hat{\sigma} \to \infty$: in that

limit, optimal rates $s^*, \alpha^* \approx 0$ for both branches, while y^* is equal to 1 in branch YA and to 0 in branch AY. Such different rates entail the same information, because we know from Eq.2.91 that the magnetization does not depend on y for $\tau = 0$.
For $\tau = 0.07$, the branch YA is still optimal for all dissipation values. However, for $\hat{\sigma} \gtrsim 3$, it matches the branch AY since the two optimal topologies coincide: $s^* \approx 0$ and $y^* \approx \alpha^* \ll 1$, that is the system starts to cycle through the four states.
For $\tau = 0.34$, branch AY is suboptimal only when $\hat{\sigma} \lesssim 2$, where rate s^* is still much larger than 0. Then it becomes optimal by featuring a perfect CW cycle where $s^* \approx y^* \approx 0$ and $\alpha^* \gtrsim 3 - 2\sqrt{2}$. On the other hand, branch YA is forced to have $y^* \geq \alpha^*$ and thus reduce information transmission in the CW cycle because of the back reaction y^*. For small dissipation values, $\mathcal{I}^*_{AY}$ is always smaller than $\mathcal{I}^*_{YA}$ because, in order to fulfill both the conditions $\alpha^* \approx s^* y^*$ and $\alpha^* \geq y^*$, the only admissible solution is to have $s^* \approx 1$. Such high value of s^* is detrimental for information transmission, as discussed in the analytical sections.

Let us conclude the discussion of Fig.2.10a with a general remark: when $\hat{\sigma} \to \infty$, we have shown that the optimal topology is a CW cycle where the system is able to transmit 1 bit of information (see section 1.4.2.4). However, when moving towards finite values of $\hat{\sigma}$, information transmission decreases, until a point where the system is confronted with a choice: either continue to cycle and reduce $\mathcal{I}$, or to concentrates on the "aligned" states and reach a finite plateau $\mathcal{I} = \mathcal{I}_{\hat{\sigma}=0}$.

We now move to the analysis of Fig.2.10b: here we display $\mathcal{I}^*$ as a function of delay τ, for different values of the rescaled dissipation $\hat{\sigma}$. Results of the optimization branches YA and AY are plotted as dotted red lines and solid cyan lines, respectively. Optimal rates used to calculate $\mathcal{I}^*$ are shown in Fig.B.2.
When $\hat{\sigma} \approx 0$ bits, the branch YA is optimal at all times, and provides a mutual information $\mathcal{I}^*_{YA}$ which decreases with τ as in Eq.2.84. Optimal rates are $s^* \approx \alpha^* \approx 0$ and $y^* = 1$. Branch AY gives always $\mathcal{I}^*_{AY} = 0$, since again $s^* = 1$ in order to fulfill $\alpha^* = s^* y^*$ and $\alpha^* \geq y^*$.
For $\hat{\sigma} = 0.23$ bits, branch AY is suboptimal until $\tau \approx 1.5$, where it matches the other branch. This happens because for $\tau \gtrsim 1.5$ the two topologies coincide: the system performs a CW cycle where $\alpha^* = y^* \approx 0.6$ and $s^* \approx 0.4$.
Finally, for very large values of dissipation ($\hat{\sigma} = 5.77$ bits), the branch AY is always optimal: for $\tau \geq (2 - \sqrt{2})$ it features the optimal topology of a CW cycle with $s^* = y^* \approx 0$ and $\alpha^* = (3 - 2\sqrt{2})$. The nontrivial decay of $\mathcal{I}^*_{AY}$ with τ was analytically derived in section 2.3.3.3. Here we just stress that the gain in information for $\tau \gtrsim (2 - \sqrt{2})$ is due to the presence of feedback, as already shown in section 1.4.3.

Optimal mutual information $\mathcal{I}^*$ is obtained as the hull curve of $\mathcal{I}^*_{YA}$ and of $\mathcal{I}^*_{AY}$. Its contour plot is shown in Fig.2.11a: we can see that, as in the circuits without feedback (Fig.2.4a), the maximum information the circuit is able to transmit decreases with the time delay of the readout for all values of $\hat{\sigma}$, as the system decorrelates with time. However, the system is now able to transmit ≈ 1 bit of information for any value of entropy dissipation, when $\tau \ll 1$.

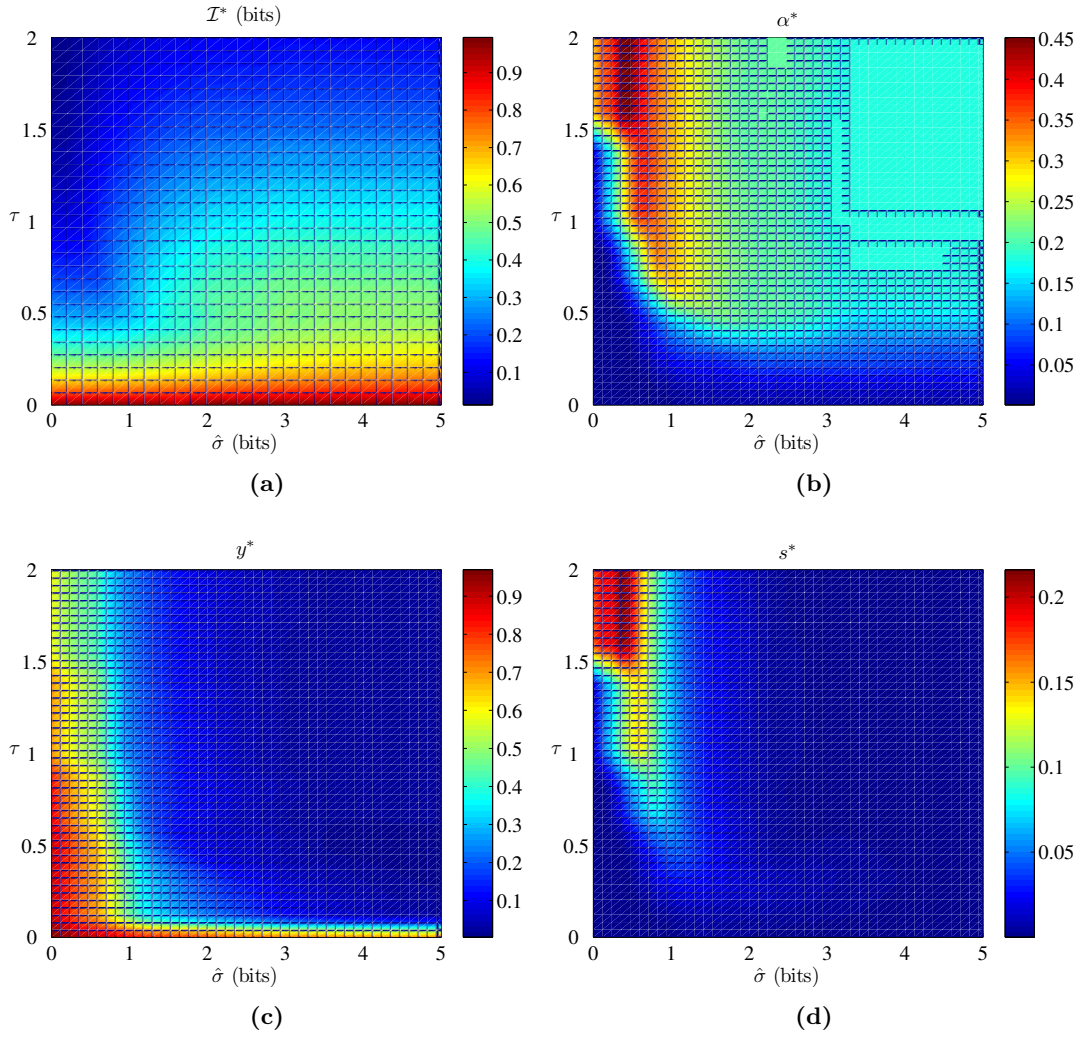
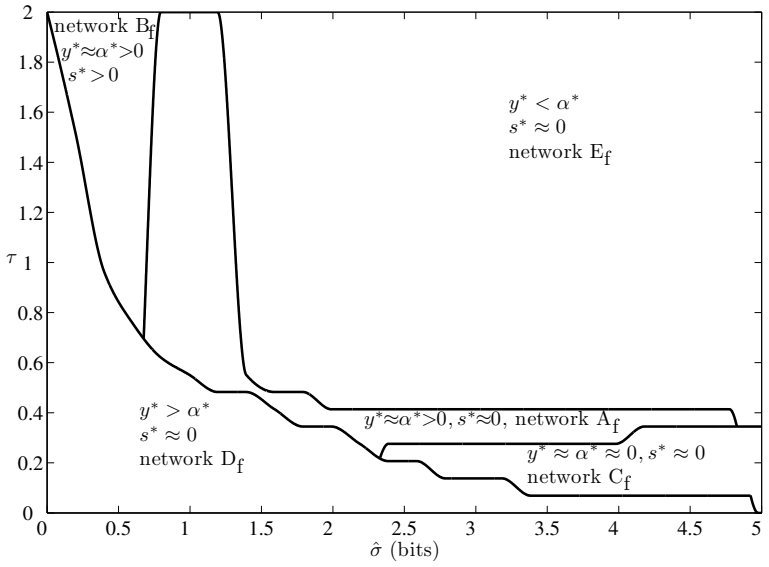


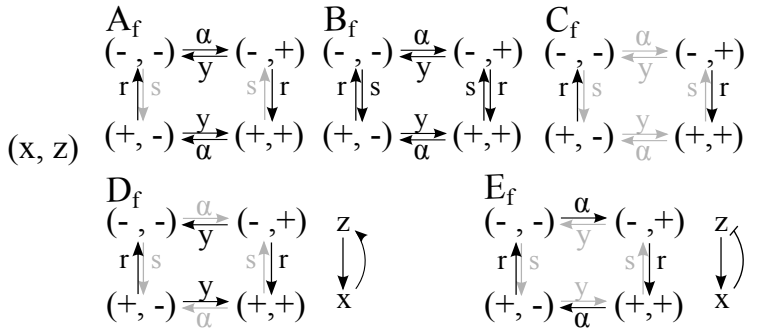
Figure 2.11: Adding feedback. (a) Contour plot of optimal mutual information $\mathcal{I}^*$ as function of the readout delay τ and entropy production rate $\hat{\sigma}$. In contrast with the simpler model of Fig.2.4a, where there is no feedback, mutual information is now equal to ≈ 1 bit for any value of $\hat{\sigma}$ when $\tau \ll 1$. (b) Contour plot of optimal input rate α^* as function of the readout delay τ and entropy production rate $\hat{\sigma}$. (c) Contour plot of optimal input rate y^* as function of the readout delay τ and entropy production rate $\hat{\sigma}$. (d) Contour plot of optimal output rate s^* as function of the readout delay τ and entropy production rate $\hat{\sigma}$.

To gain a better idea about the optimal network topologies in the presence of feedback,

we have used the rates to broadly classify the circuit topologies in the phase diagram in Fig. 2.12a with the topologies defined in Fig. 2.12b, as we did for the previous model without feedback (see section 2.3.2.1).



(a)



(b)

Figure 2.12: Adding feedback. (a) Phase diagram in the $(\hat{\sigma}, \tau)$ plane of the optimal network topologies A, B, C, D sketched in panel (b) (gray arrows stand for negligible rates).

At small but finite dissipation and large readout delays, the system has similar characteristics as the circuit with no feedback: the optimal network consists of reversible flipping of both the input and output with large rates (network B_f in Fig. 2.12b). These networks are not useful for transmitting information, but given the constraints of large time delay and close to equilibrium solutions, better topologies cannot be found. As described in section 2.3.3.1, at $\hat{\sigma} = 0$ the optimal solution has the input and output permanently fixed in the same state, providing perfect readout but not functioning as a switch. This solution is obtained with infinitely high energy barriers between the two aligned states giving infinite switching times between these two minima. Decreasing these energy barriers at small but finite dissipation results in a finite lifetime of the

two aligned states, effectively producing a switch with two stable states, encoded in the optimal motif of a positive feedback loop (network D_f in Fig. 2.12b). These optimal networks transmit close to 1 bit of information at small but finite time delays. They feature faster rates for the output and input to align, and slower rates to anti-align, resulting in larger probabilities that the system is in the aligned states at the time of the readout.

At large dissipation and small readout delays, we recover the same solution as in circuits without feedback. The input z does not change and the output quickly aligns with the input (network C_f in Fig. 2.12b). As the readout time increases, the input state switches back and forth (network A_f in Fig. 2.12b) and the system decorrelates, causing the transmitted information to decrease. At even bigger delays, the large dissipation rate allows the system to avoid the equilibrium solution of network B_f, but cycle through the states with an alternating combination of fast (r that aligns the input and output) and slow (α that anti-aligns them) rates (network E_f in Fig. 2.12b). As a result the circuit is more likely to be found in the aligned states at all times, transmitting more information. As discussed above and in the previous Chapter (section 1.4.3), the optimal network topology for large delays is a negative feedback loop, which is known to oscillate in certain parameter regimes [46]. Since oscillatory solutions would decrease the information transmitted at large delays, avoiding the oscillatory regime sets a limit to the maximum value of α. At $\hat{\sigma} < 1$ and $\tau \gtrsim 1$, when the rates antialigning the input and output increase (Figures 2.11b and 2.11d), the transmitted information decreases.

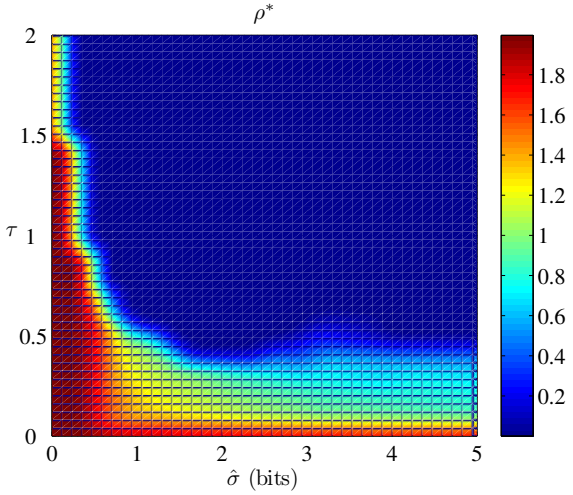


Figure 2.13: Contour plot of optimal parameter ρ^* as function of delay τ and entropy production $\hat{\sigma}$. We recall that when ρ is equal to 0 the two smallest eigenvalues of the system become degenerate and hence the two most relevant timescales match each other: in this sense, $\rho = 0$ has the same meaning of $\gamma = 1$ for the simpler model in Fig.2.4d. However, we can see here that, unlike in Fig.2.4d, the region with $\rho^* = 0$ does not reach the axis $\hat{\sigma} = 0$.

Finally, it is interesting to look at the contour plot of the optimal parameter ρ^*, shown in Fig.2.13. We remind that ρ is defined in Eq.2.66 and it is equal to 0 when the two smallest eigenvalues of the system become degenerate. This corresponds to the case with $\gamma = 1$ in the simpler model of Fig.2.4d. However, while γ^* was allowed to be 1 both for zero and infinite entropy production (see Fig.2.6), here we see that, when $\hat{\sigma} = 0$, ρ^* can only be larger than 0. We can understand this behavior intuitively if we recall the definition 2.82 for ρ at $\hat{\sigma} = 0$. It is clear that in order for ρ to be zero, both s and y must be 1 (i.e. w and v, respectively). However, looking at Figures 2.11d and 2.11c, we can easily verify that such condition is never fulfilled.

2.3.4 Comparison with the EnvZ-OmpR signaling system

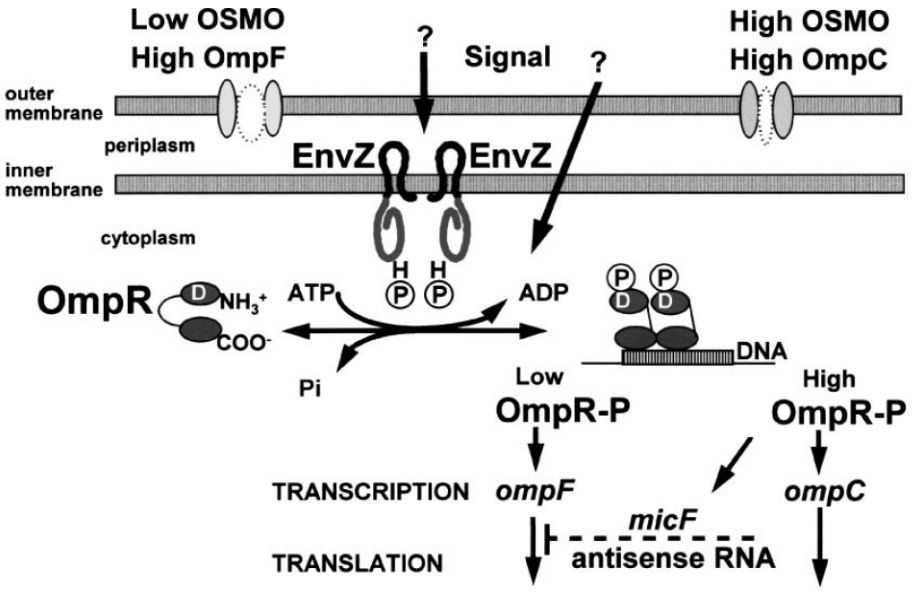


Figure 2.14: The EnvZ-OmpR phosphorelay (taken from [79]). The major components involved in osmoregulation in *E. coli* by the Histidine-Aspartate phoshorelay are shown. The OmpF and OmpC porins are trimers located in the outer membrane. EnvZ is an inner membrane histidine kinase and OmpR is a response regulator located in the cytoplasm of the cell. In response to an osmotic signal, EnvZ is activated, autophosphorylates and serves as a phosho-donor for OmpR. After OmpR is phoshorylated, there is a conformational change which allows DNA binding of OmpR-P as dimers to the OmpR-P binding sites located upstream of the porin genes *ompF* and *ompC*. Phosphorylation and dephosphorylation of OmpR is regulated by the EnvZ kinase and phosphatase activities, respectively. At low levels of osmolarity, there are lower levels of OmpR-P, resulting in the transcriptional activation of *ompF*. Conversely, at high osmolarity, there are higher levels of OmpR-P resulting in repression of *ompF* and activation of *ompC*.

Our optimal network with feedback, at large energy dissipation and relatively large readout delays (circuit E_f in Fig. 2.12b), may be related to the dynamics of the two-component signaling network in *E. coli* used in osmoregulation [79–82]. This network is composed of the histidine kinase EnvZ and the response regulator OmpR and it is aimed at reacting to an osmotic shock by regulating the expression of two porin proteins OmpF and OmpC. After phosphorylation by EnvZ, OmpR undergoes a conformational change, dimerizes and binds to the porin promoter region of either the *ompF* or *ompC* gene (see Fig.2.14). Detailed experimental studies of this system show that phosphorylation-activated dimerization drives an increase in DNA binding [82], suggesting that the biochemical regulation is a clockwise cycle such as presented in network E_f of Fig. 2.12b.

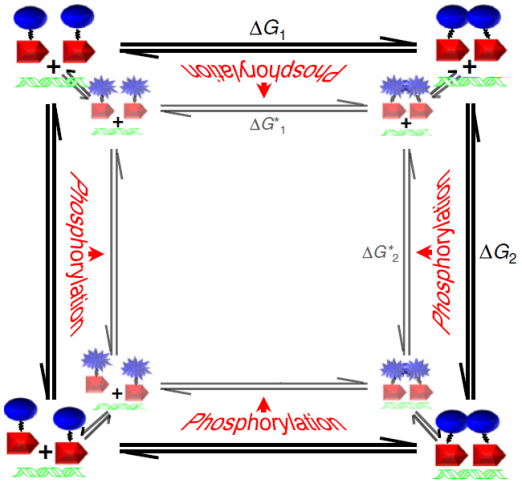


Figure 2.15: (taken from [82]) and modified. Coupled thermodynamic cycles linking OmpR dimerization and DNA binding. Blue ovals and starbursts represent OmpR receiver domains in inactive and active conformational states, respectively. Red arrows represent OmpR DNA-binding domains and DNA containing OmpR binding site sequences is depicted in green. Free energies underlying the steps in the cycle are numbered, with free energies associated with OmpR in an active conformational state being indicated by asterisks. Phosphorylation is thought to drive the conformational equilibrium toward the active state, as indicated by the red arrows pointing toward the active OmpR reaction cycle (inner cycle).

In order to better understand the thermodynamics of the process, we will consider a particular cycle in Fig.2.15. Starting from the external cycle, at the top left we have two monomers of OmpR which are phosphorylated by EnvZ and undergo a energetically favorable conformational change. Such new state is marked with starbursts at the top left in the internal cycle. The variation of Gibbs free energy $2\Delta G_{act}$ is not shown and is composed by

$$2\Delta G_{act} = 2\Delta G_{phospho} + 2\Delta G_{conform},$$

where the factor 2 stands for the presence of 2 monomers, $\Delta G_{phospho}$ is the free energy difference due to phosphorylation and $\Delta G_{conform}$ is the free energy release due to the favorable conformational change. ΔG_{act} is estimated from the literature [80] and its value reported in Table 2.1.

The two OmpR monomers then dimerize releasing a free energy ΔG_1^* (top right in the internal cycle) and then bind to either F1 or C1 DNA [3] as dimer, again releasing a free energy ΔG_2^* (bottom right in the internal cycle). The values of ΔG_1^* and ΔG_2^* are estimated from the literature [82] and listed in Table 2.1.

After binding to DNA and activating the transcription of either *ompF* or *ompC*, the OmpR is dephosphorylated and deactivated (bottom right in the external cycle) with free energy $2\Delta G_{inact}$, composed by

$$2\Delta G_{inact} = 2\Delta G_{dephospho} - 2\Delta G_{conform},$$

where the factor 2 stands again for the two dimers and $\Delta G_{dephospho}$ is the free energy release due to dephosphorylation of OmpR. ΔG_{inact} is estimated from the literature [80] and reported in Table 2.1.

Finally the dimer 2OmpR unbinds from DNA with free energy $-\Delta G_2$ (top right, external cycle) and comes back to the initial state with free energy $-\Delta G_1$ (top left, external cycle). $-\Delta G_2$ and $-\Delta G_1$ are estimated from the literature [82] and listed in Table 2.1. By using the values in Table 2.1, we can now compute the free energy released in the cycle described above, which amounts to

$$\Delta G_{tot} = 2\Delta G_{act} + (\Delta G_1^* + \Delta G_2^*) + 2\Delta G_{inact} - (\Delta G_2 + \Delta G_1) = -26.9/-59.3 \quad k_B T, \quad (2.99)$$

where the wide range of values is due to uncertainty in the energetics of the single reactions and we have considered C1 DNA as the target DNA.

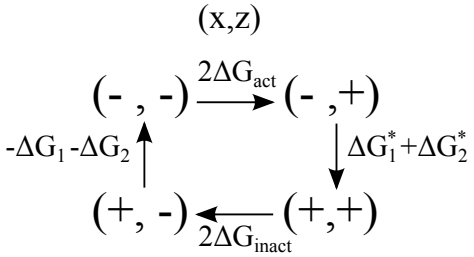


Figure 2.16: A possible matching of our four-state model with the dynamics of OmpR phophorylation, dimerization and DNA binding shown in Fig.2.15.

As shown in Fig.2.16, we can map the activation of our input z to the process of phosphorylation and conformational change of two monomers of OmpR and the activation

[3] F1 DNA corresponds to gene *ompF* and codes for porin OmpF, C1 DNA corresponds to gene *ompC* and codes for porin OmpC.

of our output x to the dimerization and binding to DNA of 2OmpR. Conversely, the deactivation of z corresponds to the dephosphorylation and deactivation of 2OmpR, while the deactivation of x corresponds to unbinding from the DNA and monomerization of 2OmpR.

We can compute the rate of heat dissipation $\dot{Q}$ as a function of the optimal information $\mathcal{I}^*$ for the network E_f of Fig. 2.12b. We first need to scale back all the quantities –rates α, s, y, r, smallest eigenvalue λ, optimal information $\mathcal{I}^*$ and entropy production rate $\hat{\sigma}$ – from absolute time τ to real time $t = \tau\lambda$ (expressed in minutes).

We thus associate rate α to the autophosphorylation rate k_{obs} of OmpR [4] taken from the literature [82], both in the presence and in the absence of C1 DNA (see Table 2.2).

	kcal/mol	k_BT	
ΔG_{act}	-5/-10	-8.1/-16.2	
ΔG_{inact}	-1/-6	-1.6/-9.7	
ΔG_1^*	-7.3	-11.9	
ΔG_1	-4.3	-7	
ΔG_2^*	-9.3	-15.1	F1 DNA
ΔG_2^*	-10.1	-16.4	C1 DNA
ΔG_2	-7.8	-12.7	F1 DNA
ΔG_2	-8.5	-13.8	C1 DNA

Table 2.1: Free energy differences ΔG for the reactions shown in Fig.2.15 [80, 82]. Error bars are not reported because these values are estimates. Conversion in k_BT units is done considering a temperature T =310 K.

$k_{obs}(min^{-1})$	
0.069	DNA n.a.
0.961	C1 DNA

Table 2.2: Autophosphorylation rate k_{obs} of OmpR [82], both in the absence and in the presence of C1 DNA. Error bars are not reported because these values are estimates.

The results are shown in Fig.2.17: in panel 2.17a we set $\tau = 1.03$, while in panel 2.17b $\tau = 2$. At a given τ we plot the dependence of the heat dissipated per minute $\dot{Q}$ on the transmitted information $\mathcal{I}^*$ for two values of the phosphorylation rate α. To transmit the same $\mathcal{I}^*$, higher phosphorylation rates entail larger amount of dissipated heat. We also note the steep increase of $\dot{Q}$ when $\mathcal{I}^*$ approaches its maximum in the allowed range: getting closer to the channel capacity requires the system to dissipate larger amounts of energy.

The final step is to compare the total heat Q_{tot} dissipated in one cycle with the total free energy released ΔG_{tot} and hence compute the network's efficiency. How do we know the duration of a cycle? It has been assessed [83] that porin proteins appear in significant quantities ≈ 10 minutes after an osmotic shock. This allows us to set such a timescale as an upper bound for the cycle duration.

[4] We chose to use the autophosphorylation rate of OmpR as a proxy of the rate of phosphorylation by the histidine kinase EnvZ, which was not available in the known literature.

So let us say that the cycle lasts ≈ 3 minutes and we are interested to know how much energy has been dissipated to optimally transmit information from the input at time $t = 0$ to the output at time $t \approx 3$ minutes. From Fig.2.17a we get $Q_{\text{tot}} \approx 18\text{k}_\text{B}\text{T}$ and $Q_{\text{tot}} \approx 36\text{k}_\text{B}\text{T}$ for transmitting information with the largest and smallest energy constraints, respectively. For a cycle lasting ≈ 6 minutes (Fig.2.17b) we would get $Q_{\text{tot}} \approx 36\text{k}_\text{B}\text{T}$ and $Q_{\text{tot}} \approx 72\text{k}_\text{B}\text{T}$, respectively. So the energetic performance of the network is strongly dependent on the cycle duration. For example, considering a cycle lasting ≈ 3 minutes, the transmission of the minimum allowed $\mathcal{I}^*$ and a $\Delta\text{G}_{\text{tot}} \approx -59.3$ $\text{k}_\text{B}\text{T}$, we may estimate the network's efficiency to be $\approx 30\%$ (see [70, 76] as a comparison).

In general, the transmission of the optimal mutual information $\mathcal{I}^*$ requires the system to dissipate into the environment an amount of heat of the same order of magnitude as the free energy released in the cycle.

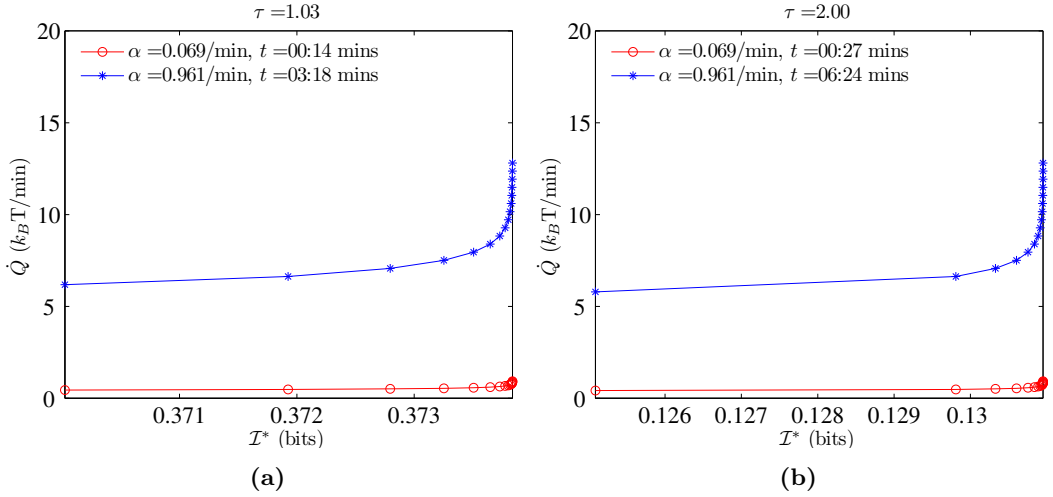


Figure 2.17: Heat dissipation rate $\dot{Q}$ as function of the maximal transmitted information $\mathcal{I}^*$, for two different absolute delays τ (panel (a) and panel (b)). Red lines with open circles and blue lines with asterisks correspond to lower and higher phosphorylation rates α, respectively (values taken from Table 2.2).

2.3.5 Discussion and conclusions

Most studies that optimize information transmission in biochemical circuits consider ideal conditions and look for the networks that are only limited by intrinsic physical constraints coming from noise in the system. However often cells must respond to signals under natural external constraints: the readout of the input occurs at a delay and cell energetics are limited. Here we investigated how these difficulties influence the form of optimal designs of biochemical circuits.

Most generally, the information transmitted by circuits decreases with the readout delay, as the system decorrelates with time. Feedback can decrease this decay, but cannot overcome it completely. In the large dissipation limit, the optimal solution consists of using a combination of fast rates for output switching and slow rates for input switching, to increase the probability of the system to be in only two states. Our choice of setting the output rate $r = 1$ fixed these two states to be the aligned states, but the natural symmetry of the system implies that a degenerate solution that transmits the same amount of information exists for the case when the input represses the output, favoring the anti-aligned states. We explicitly discussed these solution in the infinite dissipation regime in Chap.1. In the simplest circuit without feedback the only way to achieve this separation into favorable and unfavorable states is by dissipating energy and forbidding back reactions for output switching. Close to equilibrium, in the absence of feedback, the circuit cannot constrain the back reactions and, as a result, the maximum mutual information goes linearly to zero with the entropy production rate $\hat{\sigma}$ for all values of the readout delay τ. From the simplest circuit we see that the rate of input flipping depends on the time delay – longer readouts require slow flipping rates of the input to be informative, whereas the ability to dissipate energy allows the circuit to irreversible cycle through the states by eliminating both the input and output back reactions. The fully non-equilibrium solution is valid for a large range of dissipation values. If long readouts or energy constraints forbid this solution, the circuit effectively becomes randomly stuck in one of two states and not informative: the input is fixed with an equal probability to be in one of the two states, and the output attempts to align with the input. In summary, the only way for a system without feedback to transmit information is to dissipate energy.

Feedback significantly increases the range of dissipation values at which circuits can be informative. When the output feedbacks onto the input, the circuit can transmit ≈ 1 bit of information for any value of $\hat{\sigma}$ at small delays. Far from equilibrium, the optimal solution cycles through all the states, effectively increasing the decorrelation time of the system. The optimal topology is based on a negative feedback motif with a slow switching input and rapidly responding output. Such motifs are very common in stress responses (DNA damage, heat and osmotic shock and immune response) [43] and often rely on a slow (gene regulation) and fast (protein-protein interaction) step. In perfect equilibrium, the formally optimal circuit is non responsive – there is no regulation and the input and output are aligned at all times. However at small but finite entropy production rates (as well as the suboptimal solution in perfect equilibrium) the optimal topologies are different from the large dissipation case.

In the presence of feedback, the optimal circuit in the small dissipation range is a positive feedback loop with two stable states $(+,+)$ and $(-,-)$. Such circuits have long been

known to be a key mechanism for memory storage [84]. This design of a stable switch is able to convert a transient stimulus into a permanent biochemical response. These circuits have been shown to be crucial for the irreversibility of maturation of *Xenopus* oocytes [85] and for long lasting synaptic plasticity [86]. It has also been argued that positive feedback may have a role in enhancing switch-like responses (e. g. in MAP kinase cascades) and improving energetic efficiency by filtering out noise [87]. This may explain why we find such optimal topology in the small dissipation regime.

The above examples show that the optimal topology at small dissipation rates is characteristic of stable long term readouts, that commit the cell to one of two responses. The aligned (or anti-aligned in the other degenerate topology) states are very stable and large energetic barriers exist to exit these states, resulting in the positive feedback motif being optimal. Conversely, the optimal motif in the large dissipation limit is a negative feedback loop, that is characteristic of shock response – a transient response that is easily exited, but needs to be implemented quickly. It is therefore a typical non–equilibrium response, whereas the positive feedback loop is characteristic of slow and stable equilibrium situations.

Intuitively, dissipating more energy allows for larger information transmission because it lowers the probability of back reactions, which are detrimental when processing a signal. Interestingly, in the presence of feedback the system is able to build a particular topology which is suboptimal in terms of information transmission but which does not dissipate energy at all. The resulting network is such that effectively the system can cycle either in the clockwise or counterclockwise direction and the probability distribution is mostly concentrated on the aligned states $(+,+)$ and $(-,-)$. Such costless network topologies could be of inspiration when designing synthetic biochemical circuits aimed at energy production.

Feedback is able to slow down the decrease of information transmission with readout delay, but not to change the monotonic nature of this process caused by decorrelation of the states of the circuit. Yet feedback does alter the dependence of the information decay with dissipation compared to circuits without feedback. At large as well as small dissipation rates the capacity plateaus, leaving a small range of $\hat{\sigma}$ values where the transmitted information is sensitive to the precise magnitude of the energy constraints. This relatively narrow regime is where the optimal motif changes from a positive feedback loop to a negative feedback loop. Effectively in this regime the feedback is turned off (the back and forth input flipping rates are similar) and the circuit resembles the simple system discussed in section 2.3.2.

Chapter 3

Robust optimization: MaxiMin strategies

3.1 Motivation

Optimizing biochemical networks for information transmission assumes that the circuit and its environment have coevolved to best match their statistical properties. For many networks this is a valid assumption. However, other networks function in a wide variety of variable conditions. To study what kind of network is best adapted to function in adverse environments we combine a game-theoretic maximin approach with the framework of information theory. We ask what system will maximally transmit information even when presented by the environment with the worst possible initial state - the one that aims at minimizing information at all time delays. Interestingly, we find that, even if the amount of transmitted information is inevitably smaller, the structure of the optimal circuits is the same as when the environment has no detrimental effect and the system is able to optimize its initial condition.

Game-theoretic approaches have been used to robustly design biochemical networks and to devise biomimicking algorithms. Given environmental disturbances and uncertainty about the initial state, minimax strategies were used to match therapeutic treatment to a prescribed immune response [88], and to make a stochastic synthetic gene network achieve a desired steady state [89]. The adaptive response in bacterial chemotaxis has been interpreted as a maximin strategy that ensures the highest minimum chemoattractant uptake for any profile of concentration [6].

3.2 A MaxiMin model for cell regulatory networks

In many situations a biochemical circuit needs to reliably respond in many possible external conditions. In this case, optimization in the typical environment, as the one discussed in earlier sections, is not the desired criterium. Such a situation is better described by assuming that the environment chooses the worst possible conditions for the network to function. Formally this is captured by assuming that the system and the environment play a zero-sum game, where the circuit is trying to maximize the mutual information between the input and the output, while the environment is trying to minimize it.

A game theoretic formulation of the problem requires one to define the strategy space, which in this case amounts to deciding which variables are controlled by the circuit and which by the environment. Here we assume that the system will adjust the transition rates, whereas the environment controls the initial probability distribution $P_0(x_0, z_0)$ of the input z and output x.

In other words, we are interested in circuits that are optimal for working in the worst possible environmental conditions, which in game-theoretic terms correspond to maximin or "minorant" strategies [90]: the player has the goal of maximizing a function, whereas the opponent has the goal of minimizing it. This strategy is also related to "robust control" [88, 89]. In our case the circuit behaves so as to ensure that at least a certain number $\mathcal{I}$ of bits are transmitted over a given time-scale.

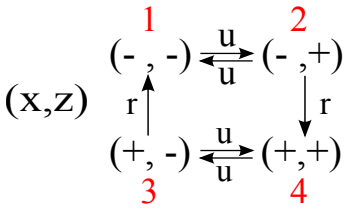


Figure 3.1: The four network states, with corresponding transition rates, considered in the maximin optimization where the input z can either up or down-regulate the output x. x aligns with z with rate r.

We look for the networks that are best adapted to the worst case scenario for the simplest circuit without feedback presented in Chap.1 and identified as model A . We recall that in this case the input z flips between the $+$ and the $-$ state with rate u, and the output aligns with the input with rate r (see Fig.3.1).

We consider this problem on the timescales of the system, which means that the system wants to maximize the mutual information $\mathcal{I}(\tau)$ between the input at time 0 and the output at a time $\tau = t\lambda$, where $\lambda = \min(r, 2u)$ is the minimal non-zero eigenvalue of the transition rate matrix $\mathcal{L}$– the inverse of the system's slowest timescale. As done in

previous Chapters, we set $r = 1$ to fix the units of time. The effective magnetization is the same as Eq.1.22:

$$\mu = \mu_0 e^{-\tau/\lambda} + \frac{1}{1-2u}\left(e^{-2u\tau/\lambda} - e^{-\tau/\lambda}\right). \tag{3.1}$$

where $\mu \geq 0$, and $|\mu_0| \leq 1$ encodes the initial condition

$$P_0(x_0, z_0) = \frac{1 + x_0 z_0 \mu_0}{4}. \tag{3.2}$$

Unlike in the cases when we optimized the transmitted information between the input and output for circuits that are in steady state, in the setup considered here the initial distribution does not need to be in steady state. The space of solutions considered here is the same as the one in section 1.4.2.5, when we optimized the information transmitted with a delay in circuits that were out of steady state. There we simultaneously found the optimal initial distribution and the parameters of the circuit. Here, we vary the same properties of the system (initial distribution and flipping rates), but with a different underlying optimization criterium – the environment minimizes the transmitted information by setting the initial distribution and the circuit sets the flipping rates.

In this specific model, the maximin game between the system and the environment takes the following form:

- The environment E chooses μ_0 so as to minimize mutual information, given the rates of the circuit. This corresponds to finding the value of μ_0 which makes μ as small as possible (since $\mathcal{I}$ is an increasing function of μ in the allowed $\mu > 1$ regime).
- Given μ_0, the circuit S looks for the rate u that maximizes $\mathcal{I}$ (i.e. μ).

The above zero-sum game between the system (circuit) S and environment E is formalized in terms of their respective cost functions $\mathcal{F}_S$ and $\mathcal{F}_E$ that satisfy

$$\mathcal{F}_S + \mathcal{F}_E = 0, \tag{3.3}$$

where $\mathcal{F}_S = -\mathcal{F}_E = |\mu| = \mathcal{F}(\mu_0, u; \tau)$. The optimization problem becomes

$$\max_u \min_{\mu_0} \mathcal{F}(\mu_0, u; \tau). \tag{3.4}$$

The optimal μ_0^* chosen by the environment is a function of u and τ, such that

$$\min_{\mu_0} \mathcal{F}(\mu_0, u; \tau) = \mathcal{F}(\mu_0^*(\tau, u), u; \tau), \tag{3.5}$$

and the circuit chooses $u^* = u^*(\tau)$ that satisfies

$$\max_u \mathcal{F}(\mu_0^*(\tau, u), u; \tau) = \mathcal{F}(\mu_0^*(\tau, u^*), u^*, \tau). \tag{3.6}$$

To make analytical progress, we have to separately consider the regime in which the smallest eigenvalue is $\lambda = 2u$ and the one in which $\lambda = 1$.

3.2.1 Case $\lambda \leq 1$

In the regime where $\lambda = 2u \leq 1$, the input switches on slower timescales then the output and the effective magnetization in Eq. 3.1 is

$$\mu(\mu_0, u; \tau) = \mu_0 e^{-\tau/2u} + \frac{1}{1-2u}\left(e^{-\tau} - e^{-\tau/2u}\right). \tag{3.7}$$

The best strategy for the environment E would be to choose μ_0^* such that $\mu = 0$. However, it is constrained to fulfill $-1 \leq \mu_0^* \leq 1$. Minimizing Eq. 3.7 with respect to u subject to the constraint on μ_0 results in:

$$\mu_0^* = \begin{cases} -\frac{e^{\tau(1-2u)/2u}-1}{1-2u}, & \tau < \tau_c(u), \\ -1, & \tau \geq \tau_c(u), \end{cases} \tag{3.8}$$

with

$$\tau_c(u) = \frac{4u}{1-2u}\log(1-u). \tag{3.9}$$

When $\tau < \tau_c(u)$, the environment is able to set μ and thus $\mathcal{I}$ to zero. However, when $\tau \geq \tau_c(u)$, the magnetization is

$$\mu(-1, u; \tau) = -e^{-\tau/2u} + \frac{1}{1-2u}\left(e^{-\tau} - e^{-\tau/2u}\right), \tag{3.10}$$

where u is constrained to be in the interval $[0, \min(1/2, u_c(\tau))]$ and $u_c(\tau)$ is obtained by inverting Eq. 3.9. Given these forms of $\mu_0^*(u, \tau)$, the circuit tries to maximize the information by tuning u at each value of τ. In the $\tau \geq \tau_c(u)$ regime the effective magnetization is maximized by a u^* that solves

$$\frac{\partial \mu(-1, u; \tau)}{\partial u}|_{u^*} = 0. \tag{3.11}$$

Eq. 3.11 results in a transcendental equation for the auxiliary variable $a = \tau\frac{1-2u}{2u}$:

$$e^{a^*\tau} = 1 + a^*\tau + 2a^{*2}\tau, \tag{3.12}$$

which needs to be solved numerically. However, in the limit $\tau \ll 1$ we can make some analytical progress and find that, when $\tau \to 0$, $u^* = \frac{\tau}{2(a^*+\tau)} \to 0$ sublinearly (see Appendix C for details of the derivation):

$$u^* \simeq \frac{\tau}{2(-\log\tau + \log(2(\log\tau)^2) - 2\frac{\log(2(\log\tau)^2)}{\log\tau})}, \tau \ll 1. \tag{3.13}$$

The above solution u^* of Eq. 3.11 is valid as long as the smallest eigenvalue $\lambda = 2u < 1$. This choice of λ constrains $u^* < 1/2$, which also constrains $\tau < \tau^*$. Setting

$$\frac{\partial\mu(-1,u;\tau)}{\partial u}|_{u^*=1/2} = 0, \tag{3.14}$$

we get the condition for $\tau^* = \tau(u^* = 1/2)$

$$\frac{1}{2}\tau^*(\tau^* - 4)e^{-\tau^*} = 0, \tag{3.15}$$

which is fulfilled by $\tau^* = 4$.

In summary, the environment E chooses μ_0^* so as to have $\mu(\mu_0^*, u; \tau) = 0$. However, this is possible only for $\tau < \tau_c(u)$. In this regime the transmitted information is always zero and there is nothing the circuit can do against the judicious choice of the environment. For $\tau \geq \tau_c(u)$ the best thing the environment E can do is to set $\mu_0^* = -1$. In order to counteract the strategy of the environment E, at each readout delay τ the circuit S chooses $u < u_c(\tau)$ (with $u_c(\tau)$ obtained by inverting Eq. 3.9), such that the environment E is forced into the regime where the best it can do is to set $\mu_0^* = -1$. In this regime, the circuit S maximizes the function $\mu(-1, u; \tau)$ in $u \in [0, \min(1/2, u_c(\tau))]$ and finds $u^* = \frac{\tau}{2(a^*+\tau)}$, where a^* is given by the solution of Eq. 3.12. The maximum value of the flipping rate for the input $u^*(\tau) = 1/2$ corresponds to the readout delay $\tau^* = 4$ and marks the transition to the regime with $\lambda = r = 1$. The effective magnetization μ^* at the transition is $3/e^4 \approx 0.05$ and hence $\mathcal{I}^* \approx 0.002$.

3.2.2 Case $\lambda = 1$

For $\tau > 4$, the smallest eigenvalue is $\lambda = r = 1$, the input switches on faster timescales than the output and the effective magnetization in Eq. 3.1 is

$$\mu = \mu_0 e^{-\tau} + \frac{1}{1-2u}\left(e^{-2u\tau} - e^{-\tau}\right). \tag{3.16}$$

The environment E chooses μ_0^* to simultaneously set $\mu = 0$ and fulfill $-1 \leq \mu_0^* \leq 1$, which gives

$$\mu_0^* = \begin{cases} -\frac{1-e^{-\tau(2u-1)}}{2u-1}, & \tau < \tau_c(u), \\ -1, & \tau \geq \tau_c(u), \end{cases} \tag{3.17}$$

with

$$\tau_c(u) = \frac{1}{2u-1} \log\left(\frac{1}{2(1-u)}\right). \tag{3.18}$$

If the system S wants to be in the regime $\tau \geq \tau_c(u)$ where $\mu_0^* = -1$, then it must choose a rate $u^* \in [1/2, u_c(\tau)]$, with $u_c(\tau)$ obtained by inverting Eq. 3.18. This choice results in the effective magnetization

$$\mu(-1, u; \tau) = \frac{2(1-u)e^{-\tau} - e^{-2u\tau}}{2u-1}. \tag{3.19}$$

For any $\tau > 4$, the effective magnetization in Eq. 3.19 is always maximum at the border $u^* = 1/2$.

In summary, for $\tau > 4$, the optimal response of the circuit is to set $u^* = 1/2$, forcing the environment into the $\tau > \tau_c$ regime where the transmitted information is larger than zero.

3.2.3 Robust Optimization Solutions

In Fig. 3.2a and 3.2b we compare the capacities and optimal input switching rates at fixed readout delay τ obtained for circuits optimized given fixed best (broken red line – results from model $\tilde{A}$ in Chap.1.4.2.5) and worst (solid blue line – the maximin strategy discussed in this section) initial conditions to the results of simply optimizing information given the system is in steady state (dotted black line – results from model A in Chap.1.4.2.2). In the first case, the environment first fixes the initial probability distribution that is most limiting (blue line) or most favorable (red line) for information transmission and the circuit then finds the switching rates that allow it to transmit most information, possibly neutralizing the harm of the environment. In the second case, the initial probability distribution is fixed at steady state and the circuit optimizes its switching rates within this constraint. We find that τ_c in Eq. 3.9 is always zero, such that the worst initial condition always corresponds to $\mu_0 = -1$ for all τ and the initial probability distribution $P_0(x_0, z_0)$ is evenly divided between the mixed states $\{(-,+),(+,-)\}$. The best initial condition has $\mu_0 = +1$ and the initial probability distribution $P_0(x_0, z_0) = 1/2$ for the aligned states $\{(+,+),(-,-)\}$. In the latter case $u^* = 1/2$ for all readout delays, and the circuit functions in a regime where the input timescale $2u$ and the output timescale $r = 1$ always match.

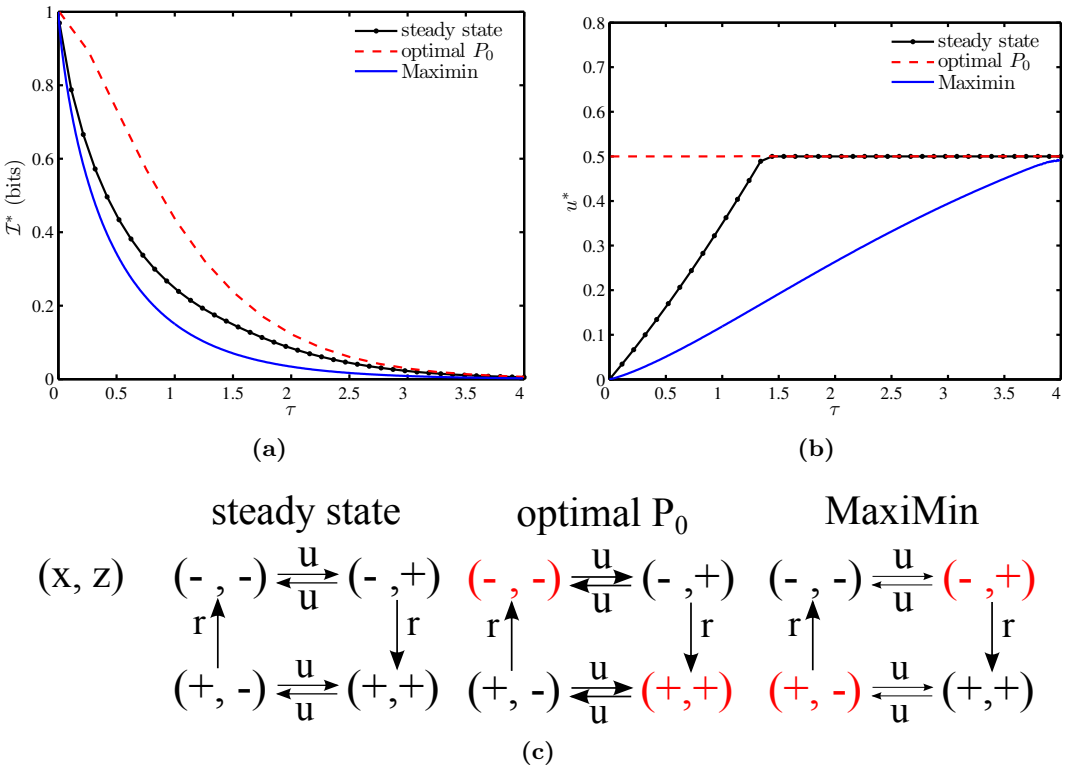


Figure 3.2: Optimal mutual information $\mathcal{I}^*$ (a) and optimal input flipping rate u^* (b) when the initial condition P_0 corresponds to the stationary state (dotted black line), is optimized by the system (dashed red line) or is set by an antagonistic environment in a maximin game (solid blue line). In panel (c) the optimal topology is shown in the three cases: states in red are the ones with initial probability $P_0 = 1/2$. Each arrow's thickness is related to the magnitude of the corresponding rate at a fixed delay $\tau = 1$.

If the initial distribution is the steady state, u^* is equal to 0 at $\tau = 0$ and increases with τ until reaching the plateau $u^* = 1/2$ for $\tau = (1 + \sqrt{3})/2$. If the environment sets the initial distribution to be the worst possible for information transmission by the circuit, $u^* = 0$ for $\tau = 0$ and increases much more slowly in τ than in the steady state circuit, finally converging to $u^* = 1/2$ for $\tau = 4$. When the circuit controls the choice of the initial state, it maximizes the probability of being in the aligned states, so that output x matches input z and the timescales of their switching are equal. However, when the environment chooses the worst initial state, forcing the initial probability distribution to be in the mixed states, the circuit requires the output x to react as fast as possible to the input z ($r \gg 2u$) to align them. Despite these differences, in all cases the optimal network takes the form of the same universal form (see Fig. 3.2c).

It is interesting to make a comparison between the effect of different initial conditions and of different energetic constraints on the choice of optimal rates: we found in Chap.2 that allowing the system to dissipate more energy postpones the *matching* of the input

and output timescales (see Fig.2.6). Likewise, as discussed above, posing the system in the worst possible initial condition, the one where "anti-aligned" states have probability 1/2, entails a larger matching time τ_c w.r.t. stationary or optimal initial conditions. A further analysis could consist in assessing the behavior of τ_c as a function of the dissipated energy and the initial conditions.

The steady-state $\mathcal{I}^*$ lies in between the optimal information in the maximin case ($\mu_0 = -1$), which we will call $\mathcal{I}^*_{\min}$, and the one where the prior is optimized ($\mu_0 = +1$), which we will indicate as $\mathcal{I}^*_{\max}$. At $\tau = 0$, all three networks transmit 1 bit of information. The maximal normalized gain $(\mathcal{I}^*_{\max} - \mathcal{I}^*_{\min})/\mathcal{I}^*_{\max}$ from optimizing the initial condition compared to the worst possible initial condition chosen by the environment has a maximum at a readout delay $\tau \approx 2.5$ (see Fig. 3.3). At this timescale, the environment can be most detrimental for information transmission.

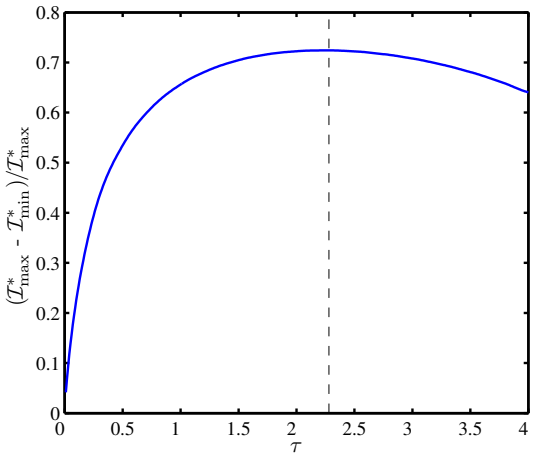


Figure 3.3: Normalized information gain $(\mathcal{I}^*_{\max} - \mathcal{I}^*_{\min})/\mathcal{I}^*_{\max}$ as function of the readout delay τ^*. Optimal $\mathcal{I}^*_{\max}$ corresponds to the case where the system optimizes the initial condition P_0, while $\mathcal{I}^*_{\min}$ corresponds to the MaxiMin solution, where the environment chooses the worst possible P_0.

3.2.4 Discussion and conclusions

In the absence of energetic constraints, we compared three different conditions in which a circuit optimizes the information transmitted at a delay for a model without feedback: a circuit that functions in steady state (model A of Chap.1.4.2.2), one that is able to optimize its input distribution (model $\tilde{A}$ of Chap.1.4.2.5), and one that is forced to function with the least informative initial distribution (maximin). Interestingly, all solutions share the same circuit topology and type of solution. The most informative solution is to cycle irreversibly through the four states. The difference between the three cases lies in the rate of flipping the input signal at a given delay. The most informative of the three strategies, where the circuit has coevolved to match the environmental conditions,

displays the largest flipping rate of the input (although still small compared to the flipping rate of the output) that is independent of the readout delay. The least informative circuit, the one that functions in an adverse environment, has the slowest flipping rate of the input. Intuitively, if the statistics of the environment and circuit match, then as long as these initial states are long lived the ability of the system to transmit information is mainly encoded in these states. However, in an adverse environment, extremely small flipping rates of the input stabilize the initial input states, allowing for a more informative readout. Since the same circuit, just with different flipping rates of the input, works optimally in both favorable and antagonistic environmental conditions, one could imagine that the rate of input switching could be tuned depending on the environmental conditions. This tuning could be achieved by fast degradation of a "typical" sugar source (like glucose) and slower degradation of a less typical sugar source (like lactose), that requires additional elements (such as production of the enzyme beta-galactosidase) for degradation.

Conclusions

In this Thesis we have applied the design principle of optimal information transmission to a simple model regulatory network with one input and one delayed output, both switching randomly between two states in continuous time. In Chapter 1 we have looked for maximally informative topologies when the system is in and out of steady state, and when feedback is present and absent. In all the setups we have identified optimal topologies which commonly occur in biology, such as the push-pull motif, typical of cellular responses to stress. For output delays larger than a certain threshold, the presence of feedback allows for more informative topologies. Moreover, when the system is not initially in steady-state but can rather optimize its initial condition, the appearance of absorbing states significantly improves the transmission of information.

In all the above setups the system was functioning strongly out of equilibrium, thus dissipating an extremely large amount of energy. In Chapter 2 we have removed such unrealistic feature by constraining the system to maximize the transmitted information while being close to thermodynamic equilibrium and consuming a fixed and finite quantity of energy. We have then discovered that at large energy dissipation optimal topologies feature negative feedback loops, typical of shock responses, while close to equilibrium positive feedback loops, known for their bistable states, become optimal. Moreover, the presence of feedback allows the system to transmit almost the maximum allowed information even in the absence of energy dissipation. A two-component osmoregulatory system of *E. coli*, composed by the kinase EnvZ and the response regulator (RR) OmpR, was taken into consideration for testing the biological soundness of the above results. The amount of energy released in a complete "cycle" of the osmoregulatory system through its four states (activation of RR by EnvZ, binding of RR to DNA, deactivation and unbinding) was roughly computed. It was estimated to be of the same order of magnitude of the heat dissipated in order to optimally transmit information.

Finally in Chapter 3 we have used a game-theoretic approach to model the interaction between the system and a detrimental environment as a maximin game. The environment chooses the initial condition so as to minimize the transmitted information at all

times, while the system tunes the input and output switching rates in order to marginalize the antagonistic effect of the environment as much as possible. We find that, unlike in the case where the initial condition is stationary or it is optimized, the system tries to make the output switch as fast as possible compared to the input. We can think of situations where the cell is presented with an unfavorable initial condition and tries to counteract it by tuning its reaction rates (e.g. "Pasteur effect").

All the above setups assume the limit of very sharp response functions, that simplify their description to two-state systems. As was previously shown, on one hand smooth regulatory functions can transmit more than 1 bit of information [5], and on the other hand the molecular noise coming from discrete particle numbers limits the capacity [12, 23–25, 33, 34, 75]. The capacity and regulatory details of the optimal systems can change if we consider more detailed molecular models. However even these simple models show general principles of how energy constraints and delayed readout drive optimal topologies. It has previously been argued using more detailed models that a truly bistable system in equilibrium is not optimal for transmitting information, unless the system does not have time to equilibrate and manages to retain memory of the initial condition [25]. The solutions we observe in our optimal networks with feedback at small dissipation correspond to circuits that manage to retain the memory of the initial state.

All the models with feedback, both in equilibrium and out of equilibrium, corresponded to two-component systems. These types of networks were previously studied as circuits that can function out of equilibrium in contrast with one-component signaling systems that must obey detailed balance [75, 91]. When it comes to precision of a continuous gradient readout, it was shown that fueling energy into the system makes it possible to overcome the limitations posed by detailed balance, by decoupling the output and receptor molecules and providing a stable readout of the input. In our discrete two-component system, this stable readout of the input state is possible even at equilibrium with a circuit design that is able to stably store the input state by exploiting timescale separation and favoring the aligned states over the non-aligned ones. However, such a stable solution is not very useful for responding to signals that change on fast timescales. In that case, energy dissipation is indispensable for an informative readout.

By studying simple two-state models of biochemical systems, we cannot interpret our optimal circuits in terms of the specific molecular designs that could be used to implement these networks. Specifically we do not account for molecular noise that comes from discrete numbers of proteins, mRNAs and genes in a regulatory circuit and that has been shown to play an important role in choosing certain regulatory elements over others [23, 29]. The role of molecular noise and cost in the design of circuits that transmit

information at a delay needs to be examined using more detailed models. The results presented here can be used as a starting point.

In our calculation we do not explicitly model fluctuations in the signal, such as was done in previous work that considered information of an instantaneous response [31] or looked at optimal delay times [41]. We just look at the optimal network that would best respond with a delay to a change in a signal. Due to this formulation we do not study fluctuations in the signal and we cannot address the question of whether the network is able to distinguish random ("irrelevant") fluctuations from a real change in the signal - a question that is very important in understanding the design of biological circuits.

This Thesis gives a framework for studying information transmission in biochemical regulatory systems subjected to delays, non-stationary conditions and energetic constraints. Such an approach can be extended to more realistic models that may explicitly account protein concentrations, more complex networks or multiple readout times.

Appendix A

Optimizing information transmission

A.1 Calculating mutual information

In Chapter 1 we calculate the mutual information between the input z at time 0 and the output x' at a time delay t, using the temporal evolution of the probability distribution $P(y)$ (with $y = (x, z)$) obtained from the master equation 1.9 [1].
The transition probability matrix $P(y'|y)$ is a solution of the master equation with initial condition

$$\lim_{t\to 0} P(y'|y) = \delta_{y',y} \tag{A.1}$$

and it can be written as the (y', y) element of the matrix $\mathrm{e}^{-t\mathcal{L}}$, i.e.

$$P(y'|y) = \left[\mathrm{e}^{-t\mathcal{L}}\right]_{y',y} = \sum_{\alpha=1}^{4} \mathrm{e}^{-\lambda_\alpha t} v_\alpha(y') u_\alpha(y). \tag{A.2}$$

λ_α (with $\alpha = 1, \ldots, 4$) are the four (assumed to be distinct for this derivation) eigenvalues of $\mathcal{L}$, and v_α and u_α^T are their corresponding orthonormal right and left eigenvectors, with components $v_\alpha(y')$ and $u_\alpha(y)$:

$$\mathcal{L} v_\alpha = \lambda_\alpha v_\alpha, \tag{A.3}$$

$$u_\alpha^T \mathcal{L} = u_\alpha^T \lambda_\alpha, \tag{A.4}$$

$$u_\alpha^T v_\beta = \delta_{\alpha,\beta}. \tag{A.5}$$

[1] The implied dependence on time is omitted, primed variables (e.g., y') refer to the state of the system at time $t \neq 0$ and unprimed ones refer to the state at $t = 0$.

In particular, if we choose a normalization such that $u_1 = (1, 1, 1, 1)$, the eigenvector v_1, corresponding to the eigenvalue $\lambda_1 = 0$, is the stationary state $P_\infty(y)$.

In order to compute the mutual information $I[x_t, z_0]$, as defined in Eq. 1.11, we need to derive an expression for the joint probability $P(x', z)$ in terms of conditional probabilities:

$$\begin{aligned} P(x', z) &= \\ &= \sum_{y,y'} P(x', z|y', y)P(y', y) \quad \text{using the definition of conditional probabilities} \\ &= \sum_{y,y'} P(x'|y')P(z|y)P(y', y) \quad \text{exploiting the conditional independence of } x', z \\ &= \sum_{y,y'} P(x'|y')P(z|y)P(y'|y)P_0(y) \quad \text{using the definition of conditional probabilities.} \end{aligned} \tag{A.6}$$

Note that the elements of $P(x'|y')$ and $P(z|y)$ are either 0 or 1 according to whether, for example, y is consistent or inconsistent with z:

$$\begin{aligned} P(z = +|y = (+, +)) &= 1 \\ P(z = +|y = (+, -)) &= 0 \\ &\vdots \end{aligned}$$

et cetera. Finally, the marginal probabilities $P(z)$ and $P(x')$ are given by

$$P(z) \equiv \sum_{x'} P(x', z), \quad P(x') \equiv \sum_{z} P(x', z).$$

The numerical computation of the mutual information can now be implemented and the optimal rates for systems of various complexity can be found numerically. In the paragraphs below we present useful computational details for implementing this calculation. We have implemented the optimization procedure in MATLAB and we made the source code available via the following public access repository: http://infodyn.sourceforge.net. For certain models we can also make analytical progress by exploiting spectral representations of the joint distribution, as shown in Chapter 1.

A.1.1 Numerical computation of the joint distribution

For numerical computation in MATLAB, it is useful to rewrite Eq. A.6 in terms of matrix operations. To that end, we define (note that X and Z are $0 - 1$ matrices –

whose elements are 0 or 1, as per Eq. A.7)

$$X_{x',y'} \equiv P(x'|y') \tag{A.7}$$
$$G_{y',y_0} \equiv P(y'|y_0) = \left[\mathrm{e}^{-t\mathcal{L}}\right]_{y',y_0} \tag{A.8}$$
$$P_{y_0,y} \equiv P_0(y_0)\delta_{y_0,y} \tag{A.9}$$
$$Z_{y,z} \equiv P(z|y). \tag{A.10}$$

This allows us to write Eq. A.6 compactly as

$$P(x',z) = \sum_{y'} X_{x,y'} \sum_{y_0} G_{y',y_0} \sum_{y} P_{y_0,y} Z_{y,z} \tag{A.11}$$
$$= [XGPZ]_{x',z} \tag{A.12}$$

that is how the equation is implemented in MATLAB.

A.1.2 Analytical calculation of the joint distribution

For analytic calculations, it is useful to expand $P(y'|y)$ in Eq. A.6 in terms of its spectral representation (Eq. A.2):

$$\begin{aligned} P(x',z) &= \sum_{\alpha}\sum_{y',y} P(x'|y')P(z|y)\mathrm{e}^{-\lambda_\alpha t} v_\alpha(y') u_\alpha(y) P_0(y) \\ &= \sum_{\alpha} \mathrm{e}^{-\lambda_\alpha t} \left(\sum_{y'} P(x'|y') v_\alpha(y')\right)\left(\sum_{y} P(z|y) u_\alpha(y) P_0(y)\right) \\ &= \left(\sum_{y'} P(x'|y') P_\infty(y')\right)\left(\sum_{y} P(z|y) P_0(y)\right) + \\ &+ \sum_{\alpha>1} \mathrm{e}^{-\lambda_\alpha t} \left(\sum_{y'} P(x'|y') v_\alpha(y')\right)\left(\sum_{y} P(z|y) u_\alpha(y) P_0(y)\right) \\ &\equiv P_\infty(x') P_0(z) + \sum_{\alpha>0} \mathrm{e}^{-\lambda_\alpha t} \tilde{v}_\alpha^{x'} \tilde{u}_\alpha^{z} \end{aligned} \tag{A.13}$$

where

$$P_\infty(x') \equiv \sum_{y'} P(x'|y')P_\infty(y') \tag{A.14}$$

$$P_0(z) \equiv \sum_{y} P(z|y)P_0(y) \tag{A.15}$$

$$\tilde{v}_\alpha^{x'} \equiv \sum_{y'} v(y')_\alpha P(x'|y') \tag{A.16}$$

$$\tilde{u}_\alpha^{z} \equiv \sum_{y} u_\alpha(y)P_0(y)P(z|y). \tag{A.17}$$

Writing $P(x', z)$ in this form makes it clear that, if the eigenvalues are distinct and thus $P(y'|y)$ is diagonalizable, then $P(x', z)$ factorizes as $t \to \infty$ and thus $I[x_t, z_0] \to 0$. Also clear is that $P(x', z) - P_\infty(x')P_0(z)$ is expressible as a sum of time-decaying exponentials. Since $P(x'|y')$ and $P(z|y)$ are $0-1$ matrices, in many cases $\tilde{v}_\alpha$ and $\tilde{u}_\alpha$ can be calculated explicitly, as shown below.

A.2 Optimization procedure

We optimize over the parameters of each model in order to maximize $\mathcal{I}(\tau) = I[x_{t=\tau/\lambda}, z_0]$, where τ is a dimensionless quantity that results from the rescaling procedure:

$$t \to t \cdot \lambda \equiv \tau,$$

where λ is the inverse of the system's largest relaxation rate (the smallest nonzero eigenvalue of the rate matrix $\mathcal{L}$). The steps of the "rescale and optimize" procedure are:

while $\tau_{min} < \tau < \tau_{max}$:

1. optimize $\mathcal{I}(\tau; \theta)$ over parameters θ or parameters θ and initial distribution $P_0(y)$
2. save $\mathcal{I}^*, \theta^*, P_0^*$
3. increment τ

end loop over τ

where

calculate $\mathcal{I}(\tau; \theta)$:

1. calculate $\mathcal{L}(\theta)$

2. calculate $\lambda(\mathcal{L})$

3. calculate $P(x,z) = X\exp(-\tau\mathcal{L}/\lambda)PZ$, as in Eq. A.12

4. calculate $I[P(x,z)]$

   return $I[P(x,z)]$ to optimization algorithm

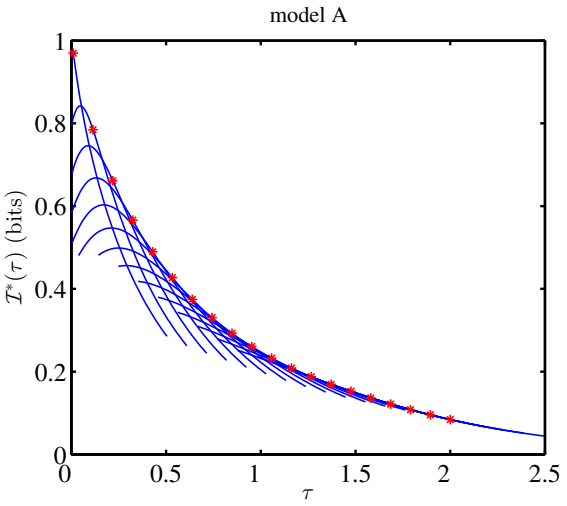


Figure A.1: An explicit construction of the optimal information curves presented in Chapter 1, here shown for model A . The maximum value $\mathcal{I}^*(\tau)$ for each τ (red $*$) is obtained by optimizing the rates u and r at each τ. The optimal rates can be different for each τ. The blue continuous curves show the whole range of $\mathcal{I}(\tau)$: each of them intersects a point $(\tau_{\mathcal{I}^*}, \mathcal{I}^*(\tau))$ and is computed by using the corresponding optimal set of rates.

The results are obtained as hull plots, as presented in Fig. A.1 for model A .

Appendix B

The energetic cost of information transmission

B.1 Simplest model

B.1.1 Limit $\tau \ll 1$

Here we detail the solutions of the equations presented in the small τ limit of the simplest model in section 2.3.2. In the small dissipation limit we assume

$$\gamma^* \simeq \frac{a_0}{\tau} + b_0 + c_0\tau \tag{B.1}$$

and find the coefficients of the expansion by solving $\frac{d\mu}{d\gamma}|_{\gamma^*} = 0$ (with μ given by Eq. 2.39) order by order in τ. The coefficient $a_0 = 0.96...$ is given by the solution of the transcendental equation $e^{a_0} = \frac{4}{3}(a_0 + 1)$, while b_0 and c_0 are given respectively by

$$\begin{aligned} b_0 &= 1 + \frac{-5 - 2a_0}{6a_0^2} = -0.25... \quad \text{and} \\ c_0 &= \frac{-75 - 65a_0 + 128a_0^2 + 28a_0^3}{72a_0^5} = 0.1... \end{aligned}$$

Using Eq. B.1 for γ^* gives

$$\mu^* \simeq \sqrt{\hat{\sigma}}(1 + A_0\tau + B_0\tau^2), \tag{B.2}$$

with

$$\begin{aligned} A_0 &= \frac{-4e^{-a_0}+3}{2a_0} - 1 = -0.24... \quad \text{and} \\ B_0 &= e^{-a_0}\frac{-5-7a_0+a_0^2+6a_0^3}{3a_0^4} + \frac{10+4a_0-a_0^2-12a_0^3+4a_0^4}{8a_0^4} = 0.01... \end{aligned}$$

Similarly, in the large dissipation limit we take

$$\gamma^* \simeq \frac{a_\infty}{\tau} + b_\infty + c_\infty \tau \tag{B.3}$$

and following the same procedure as above with μ given by Eq. 2.43 we find $a_\infty = 1.68...$ as the solution of the transcendental equation $e^{a_\infty} = 2(a_\infty + 1)$, while b_∞ and c_∞ are given by

$$\begin{aligned} b_\infty &= 1 + \frac{-2-a_\infty}{a_\infty^2} = -0.31... \quad \text{and} \\ c_\infty &= \frac{-12-12a_\infty+7a_\infty^2+3a_\infty^3}{2a_\infty^5} = 0.07... \end{aligned}$$

The effective magnetization is

$$\mu^* \simeq 1 + A_\infty \tau + B_\infty \tau^2, \tag{B.4}$$

with

$$\begin{aligned} A_\infty &= \frac{-2e^{-a_\infty}+1}{a_\infty} - 1 = -0.63... \quad \text{and} \\ B_\infty &= \frac{1}{2} + \frac{(e^{-a_\infty}(-4-6a_\infty+2a_\infty^2)+2+a_\infty-a_\infty^3)}{a_\infty^4} = 0.23... \end{aligned}$$

B.2 Adding feedback

B.2.1 Computing μ

In order to compute the effective magnetization μ in the presence of feedback, we recall Eq. 2.25, which is valid in general and which relates the joint probability distribution $P(x_t, z_0)$ with μ. We then express $P(x_t, z_0)$ as in Eq.2.13 in the main text:

$$P(x_t, z_0) = \sum_{z_t, x_0 = \pm 1} P(x_t, z_t, t | x_0, z_0, 0) P(x_0, z_0), \tag{B.5}$$

where $P(x_0, z_0)$ is the initial distribution of the system, corresponding to the stationary state $P_\infty \equiv v_1(x_0, z_0)$, while the conditional probability $P(x_t, z_t, t|x_0, z_0, 0)$ can be

written as

$$P(x_t, z_t, t|x_0, z_0, 0) = \sum_{i=1}^{4} e^{-\lambda_i t} u_i^T(x_0, z_0) v_i(x_t, z_t). \tag{B.6}$$

v_i denotes the i-th right eigenvector and u_i^T – the i-th left eigenvector and we make the dependence on x and z explicit as we are going to exploit it in the subsequent algebraic manipulations.

We recall the definitions of A and ρ

$$A = 1 + s + y + \alpha \tag{B.7}$$
$$\rho = \sqrt{(1 + s + y + \alpha)^2 - 8(sy + \alpha)}, \tag{B.8}$$

and introduce the additional quantities

$$q = \frac{1 + y - s - \alpha}{A}, \tag{B.9}$$
$$m = \frac{s + \alpha}{2A}. \tag{B.10}$$

We rewrite the right eigenvectors of Equations (2.68,2.69,2.70,2.71) as

$$v_1(x, z) = \frac{1 + qxz}{4} \tag{B.11}$$
$$v_2(x, z) = mxz \tag{B.12}$$
$$v_3(x, z) = -\tfrac{1+y-3k+\alpha+\rho}{8\rho} x - \tfrac{1+y+s-3\alpha+\rho}{8\rho} z \tag{B.13}$$
$$v_4(x, z) = \tfrac{1+y-3k+\alpha-\rho}{8\rho} x + \tfrac{1+y+s-3\alpha-\rho}{8\rho} z \tag{B.14}$$

and define

$$h = \frac{1 + y}{s + \alpha}, \tag{B.15}$$
$$a = \frac{1}{2}(1 + s - 3y + \alpha - \rho), \tag{B.16}$$
$$b = \frac{1}{2}(-3 + s + y + \alpha - \rho), \tag{B.17}$$
$$c = 1 - s + y - \alpha + \rho, \tag{B.18}$$
$$e = \frac{1}{2}(1 + s - 3y + \alpha + \rho), \tag{B.19}$$
$$f = \frac{1}{2}(-3 + s + y + \alpha + \rho), \tag{B.20}$$
$$g = 1 - s + y - \alpha - \rho. \tag{B.21}$$

Having done that, the left eigenvectors of Equations (2.72,2.73,2.74,2.75) now read

$$u_1^T(x,z) = 1, \tag{B.22}$$
$$u_2^T(x,z) = \frac{1-h}{2} + \frac{1+h}{2}xz, \tag{B.23}$$
$$u_3^T(x,z) = \frac{ax+bz}{c} \tag{B.24}$$
$$u_4^T(x,z) = \frac{ex+fz}{g}. \tag{B.25}$$

Now, by plugging Eq. B.6 into Eq. B.5, we are able to write $P(x_t, z_0)$ as

$$P(x_t, z_0) = \sum_{i=1}^{4} e^{-\lambda_i t} A_i(z_0) B_i(x_t), \tag{B.26}$$

where

$$A_i(z_0) = \sum_{x_0 \pm 1} u_i^T(x_0, z_0) v_1(x_0, z_0), \tag{B.27}$$
$$B_i(x_t) = \sum_{z_t \pm 1} v_i(x_t, z_t). \tag{B.28}$$

Computing the terms A_i and B_i (with $i = 1, \ldots, 4$) we obtain:

$$A_1(z_0) = 1/2, \tag{B.29}$$
$$A_2(z_0) = \frac{1}{4}\left(1 - h + (1+h)qz_0^2\right), \tag{B.30}$$
$$A_3(z_0) = \frac{(b+aq)z_0}{2c}, \tag{B.31}$$
$$A_4(z_0) = \frac{(f+eq)z_0}{2g}, \tag{B.32}$$

and

$$B_1(x_t) = 1/2, \tag{B.33}$$
$$B_2(x_t) = 0, \tag{B.34}$$
$$B_3(x_t) = -\frac{x_t(1-3s+y+\alpha+\rho)}{4\rho}, \tag{B.35}$$
$$B_4(x_t) = \frac{x_t(1-3s+y+\alpha-\rho)}{4\rho}. \tag{B.36}$$

Plugging in all the above expressions into Eq. B.26 we compute the effective magnetization μ, which is

$$\mu = \exp\left(-\frac{A}{2\lambda}\tau\right)\left\{q\cosh\left(\frac{\rho}{2\lambda}\tau\right) - \frac{\left[s^2-(1+y)^2-4\alpha+\alpha^2+2s(2y+\alpha)\right]}{A\rho}\sinh\left(\frac{\rho}{2\lambda}\tau\right)\right\}. \tag{B.37}$$

B.2.2 Numerical results: optimal rates

In this Appendix we show the optimal rates $\{s^*, y^*, \alpha^*\}$ resulting from numerical optimization in the model with feedback of Chapter 2. As discussed in the main text, optimization is performed as two separates branches: one where we fix $y \geq \alpha$, and the other one where we set $y \leq \alpha$. Results from both branches are shown in Figs. B.1 and B.2.

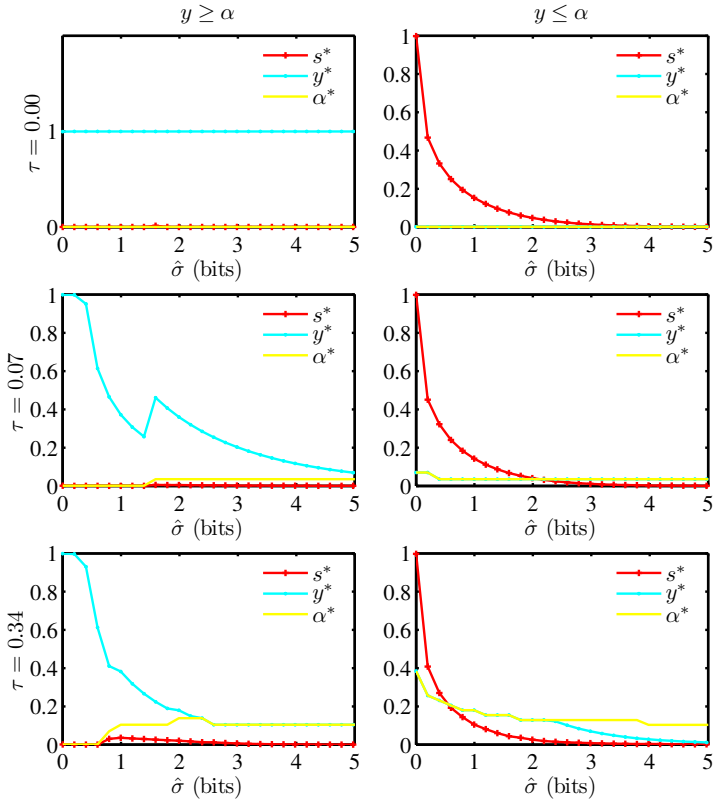


Figure B.1: Optimal rates $\{s^*, y^*, \alpha^*\}$ as functions of rescaled dissipation $\hat{\sigma}$. Values for $\tau = 0$, $\tau = 0.07$ and $\tau = 0.34$ are shown at the top, center and bottom, respectively. Results from the simulation branch with $y \geq \alpha$ and with $y \leq \alpha$ are displayed on the left and on the right, respectively. These rates are used to compute the optimal mutual information $\mathcal{I}^*$ of Fig.2.10a.

In Fig. B.1, we show the optimal rates as functions of rescaled dissipation $\hat{\sigma}$, for different values of delay τ. Such rates are used to calculate the optimal mutual information shown in Fig. 2.10a in the main text.

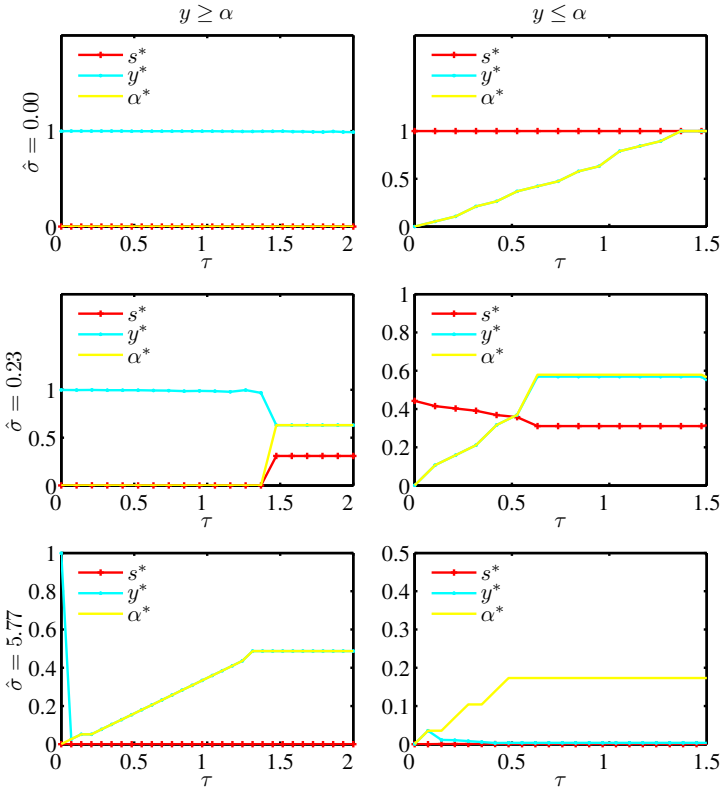


Figure B.2: Optimal rates $\{s^*, y^*, \alpha^*\}$ as functions of delay τ. Values for $\hat{\sigma} = 0$, $\hat{\sigma} = 0.23$ bits and $\hat{\sigma} = 5.77$ bits are shown at the top, center and bottom, respectively. Results from the simulation branch with $y \geq \alpha$ and with $y \leq \alpha$ are displayed on the left and on the right, respectively. These rates are used to compute the optimal mutual information $\mathcal{I}^*$ of Fig.2.10b.

In Fig. B.2 we show the dependency of the optimal rates τ, for different values of $\hat{\sigma}$. These corresponds to the maximal mutual information $\mathcal{I}^*$ shown in Fig. 2.10b in the main text.

Appendix C

Robust optimization: MaxiMin strategies

In this Appendix we derive the asymptotic behavior of u^* as $\tau \to 0$ given in Eq. 3.13. In general, the condition $\frac{\partial \mu(-1,u;\tau)}{\partial u}|_{u^*} = 0$ in Eq. 3.11 results in a transcendental equation for the auxiliary variable $a = \tau \frac{1-2u}{2u}$:

$$e^{a^*} = 1 + a^* + \frac{2a^{*2}}{\tau}, \tag{C.1}$$

which needs to be solved numerically. However, in the limit $\tau \ll 1$ we can solve Eq. C.1 analytically.

When $\tau \to 0$, in order for the left hand side (l.h.s.) of Eq. C.1 to match the leading order term $1/\tau$ of the right hand side (r.h.s.) of Eq. C.1 , a^* must be of the form

$$a^* = -\log \tau + b. \tag{C.2}$$

Hence Eq. C.1 becomes

$$\frac{e^b}{\tau} = 1 + b - \log \tau + \frac{1}{\tau} 2(b - \log \tau)^2. \tag{C.3}$$

Multiplying both sides by τ gives

$$e^b = (1 + b - \log \tau)\tau + 2b^2 - 4b \log \tau + 2(\log \tau)^2. \tag{C.4}$$

The leading order term of the r.h.s. for $\tau \to 0$ is $2(\log \tau)^2$, thus the above equation becomes

$$e^b \simeq 2(\log \tau)^2, \quad \tau \to 0,$$

which implies that b has the form

$$b = \log(2(\log\tau)^2) + c. \tag{C.5}$$

Plugging b into Equation C.4 we obtain

$$\begin{aligned}(2(\log\tau)^2)e^c &\simeq 2(\log\tau)^2 - 4(\log(2(\log\tau)^2) + c)\log\tau + \\ &+ 2(\log(2(\log\tau)^2) + c)^2 + \dots .\end{aligned} \tag{C.6}$$

and dividing by $2(\log\tau)^2$, we have

$$e^c = 1 - 2\frac{\log(2(\log\tau)^2) + c}{\log\tau} + \dots \simeq e^{-2\frac{\log(2(\log\tau)^2)}{\log\tau}},$$

which finally implies that

$$c = -2\frac{\log(2(\log\tau)^2)}{\log\tau}. \tag{C.7}$$

To sum up, when $\tau \ll 1$ one can write a^* as

$$\begin{aligned}a^* &\simeq -\log\tau + \log(2(\log\tau)^2) - 2\tfrac{\log(2(\log\tau)^2)}{\log\tau} \\ &= \log\left(\tfrac{2(\log\tau)^2}{\tau}\right) - 2\tfrac{\log(2(\log\tau)^2)}{\log\tau}\end{aligned} \tag{C.8}$$

When $\tau \to 0$, a^* diverges as

$$a^* \simeq \log\left(\frac{2(\log\tau)^2}{\tau}\right) - 2\frac{\log(2(\log\tau)^2)}{\log\tau}, \quad \tau \ll 1 \tag{C.9}$$

and u^* goes to zero with τ in a strongly sublinear way:

$$u^* \simeq \frac{\tau}{2(-\log\tau + \log(2(\log\tau)^2) - 2\frac{\log(2(\log\tau)^2)}{\log\tau})}, \tau \ll 1. \tag{C.10}$$

Bibliography

[1] William Bialek. *Biophysics: Searching for Principles*. Princeton University Press, 2012.

[2] Paul François, Vincent Hakim, and Eric D Siggia. Deriving structure from evolution: metazoan segmentation. *Molecular Systems Biology*, 3(1), 2007. ISSN 1744-4292.

[3] Ulrich Gerland and Terence Hwa. Evolutionary selection between alternative modes of gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(22):8841–6, June 2009. ISSN 1091-6490.

[4] Filipe Tostevin, Pieter Rein ten Wolde, and Martin Howard. Fundamental limits to position determination by concentration gradients. *PLoS computational biology*, 3(4):e78, April 2007. ISSN 1553-7358.

[5] Gašper Tkačik, Curtis G. Callan Junior, and William Bialek. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(34):12265–12270, August 2008.

[6] Antonio Celani and Massimo Vergassola. Bacterial strategies for chemotaxis response. *Proceedings of the National Academy of Sciences of the United States of America*, 107(4):1391–1396, 2010.

[7] Pankaj Mehta, Sidhartha Goyal, Tao Long, Bonnie L Bassler, and Ned S Wingreen. Information processing and signal integration in bacterial quorum sensing. *Molecular systems biology*, 5(325):325, January 2009. ISSN 1744-4292.

[8] Timothy Saunders and Martin Howard. When it pays to rush: interpreting morphogen gradients prior to steady-state. *Physical Biology*, 6:046020, 2009.

[9] J Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences of the United States of America*, 71(10):4135–4139, 1974.

[10] Paul François and Eric D Siggia. Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development (Cambridge, England)*, 137 (14):2385–95, July 2010. ISSN 1477-9129.

[11] M.A. Savageau. Design of molecular control mechanisms and the demand for gene expression. *Proceedings of the National Academy of Sciences*, 74(12):5647–5651, December 1977. ISSN 0027-8424.

[12] Gašper Tkačik and Aleksandra M. Walczak. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter*, 23:153102, 2011.

[13] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[14] Peter A. Lawrence. *The Making of a Fly : The Genetics of Animal Design*. Blackwell Scientific Publications, Oxford, 1992.

[15] Aude Porcher and Nathalie Dostatni. The bicoid morphogen system. *Current Biology*, 20:R249–R254, 2010.

[16] Thomas Gregor, Eric F. Wieschaus, Alistair P. McGregor, William Bialek, and David W. Tank. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell*, 130(1):141–152, July 2007.

[17] Thomas Gregor, David W. Tank, Eric F. Wieschaus, and William Bialek. Probing the limits to positional information. *Cell*, 130(1):153–164, July 2007.

[18] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.

[19] E T Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[20] Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8:450, 2007.

[21] S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon. The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *Journal of Molecular Biology*, 356(5):1073 – 1081, 2006. ISSN 0022-2836.

[22] Eldon Emberly. Optimizing the readout of morphogen gradients. *Phys. Rev. E*, 77: 041903, Apr 2008.

[23] G. Tkačik, A. M. Walczak, and W. Bialek. Optimizing information flow in small genetic networks. *Physical Review E*, 80:031920, 2009.

[24] A.M. Walczak, G. Tkačik, and W. Bialek. Optimizing information flow in small genetic networks. ii. feed-forward interactions. *Physical Review E*, 81:041905, 2010.

[25] G. Tkačik, A. M. Walczak, and W. Bialek. Optimizing information flow in small genetic networks. iii. a self-interacting gene. *Physical Review E*, 85:041903, 2012.

[26] Raymond Cheong, Alex Rhee, Chiaochun Joanne Wang, Ilya Nemenman, and Andre Levchenko. Information transduction capacity of noisy biochemical signaling. *Science*, 334(354):354, October 2011.

[27] Jangir Selimkhanov, Brooks Taylor, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev Tsimring, and Roy Wollman. Accurate information transmission through dynamic biochemical signaling networks. *Science*, 346(6215):1370–1373, 2014.

[28] Jürgen Pahle, Anne K Green, C Jane Dixon, and Ursula Kummer. Information transfer in signaling pathways: A study using coupled simulated and experimental data. *BMC Bioinformatics*, 9:139, March 2008.

[29] Aleksandra M. Walczak, Andrew Mugler, and Chris H. Wiggins. A stochastic spectral analysis of transcriptional regulatory cascades. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6529–6534, April 2009.

[30] A. Mugler, A.M. Walczak, and C.H. Wiggins. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Physical Review E*, 80:041921, 2009.

[31] Joseph Levine, Hao Yuan Kueh, and Leonid Mirny. Intrinsic fluctuations, robustness, and tunability in signaling cycles. *Biophysical Journal*, 92:4473–4481, 2007.

[32] Andrew Mugler, Aleksandra M. Walczak, , and Chris H. Wiggins. Information-optimal transcriptional response to oscillatory driving. *Physical Review Letters*, 105:058101, 2010.

[33] Filipe Tostevin and Pieter Rein ten Wolde. Mutual information between input and output trajectories of biochemical networks. *Physical Review Letters*, 102:218101, May 2009.

[34] Filipe Tostevin and Pieter Rein ten Wolde. Mutual information in time-varying biochemical systems. *Physical Review E*, 81:061917, 2010.

[35] Wiet Hendrik de Ronde, Filipe Tostevin, and Pieter Rein ten Wolde. Effect of feedback on the fidelity of information transmission of time-varying signals. *Physical Review E*, 82:031914, 2010.

[36] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde. Feed-forward loops and diamond motifs lead to tunable transmission of information in the frequency domain. *Physical Review E*, 86:021913, 2012.

[37] Alberts, Bray, Hopkin, Johnson, Lewis, Raff, Roberts, and Walter. *Essential Cell Biology*. Garland Science, New York, 3rd edition, 2009.

[38] Jeffrey E. Segall, Steven M. Block, and Howard C. Berg. Temporal comparisons in bacterial chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 83:8987–8991, December 1986.

[39] D. Fuller, W. Chen, M. Adler, A. Groisman, H. Levine, W.-J. Rappel, and W. F. Loomis. External and internal constraints on eukaryotic chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 107:9656, 2010.

[40] Thuc T. Le, Thierry Emonet, Sebastien Harlepp, Caĭin C. Guet, and Philippe Cluzel. Dynamical determinants of drug-inducible gene expression in a single bacterium. *Biophysical Journal*, 90:3315–3321, May 2006.

[41] Ilya Nemenman. Gain control in molecular information processing: Lessons from neuroscience. *Phys. Biol.*, 9:026003, 2012.

[42] Eric Guisbert, Christophe Herman, Chi Zen Lu, and Carol A. Gross. A chaperone network controls the heat shock response in e. coli. *Genes & Development*, 18: 2812–2821, 2004.

[43] Galit Lahav, Nitzan Rosenfeld, Alex Sigal, Naama Geva-Zatorsky, Arnold J Levine, Michael B Elowitz, and Uri Alon. Dynamics of the p53-mdm2 feedback loop in individual cells. *Nature Genetics*, 36(2):147, February 2004.

[44] Allan R. Brasier. The nf-κ b regulatory network. *Cardiovascular Toxicology*, 6: 111–130, 2006.

[45] Olivier Pourquié. The segmentation clock: Converting embryonic time into spatial pattern. *Science*, 301:328, 2003.

[46] Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature (London)*, 456:516, 2008.

[47] Simon Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsh*, 36 c:910–912, 1981.

[48] Lindsey D. Mayo and David B. Donner. The pten, mdm2, p53 tumor suppressor-oncoprotein network. *Trends in Biochemical Sciences*, 27:462–467, 2002.

[49] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75:126001, 2012.

[50] György Szabó, Tâniá Tomé, and István Borsos. Probability currents and entropy production in nonequilibrium lattice systems. *Physical Review E*, 82:011105, 2010.

[51] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern Physics*, 48(4), 1976.

[52] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, 2006.

[53] J Ninio. Kinetic amplification of enzyme discrimination. *Biochimie*, 57(5):587–95, 1975.

[54] Filipe Tostevin and Martin Howard. A stochastic model of min oscillations in escherichia coli and min protein segregation during cell division. *Physical Biology*, 3(1):1, 2006.

[55] Filipe Tostevin and Martin Howard. Modeling the establishment of par protein polarity in the one-cell c. elegans embryo. *Biophysical Journal*, 95(10):4512 – 4522, 2008. ISSN 0006-3495.

[56] P François and Vincent Hakim. Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19):580–584, 2004.

[57] G. Tkačik, T. Gregor, and W. Bialek. The role of input noise in transcriptional regulation. *PLoS ONE*, 3:e2774, 2008.

[58] Julien O. Dubuis, Gašper Tkačik, Eric F. Wieschaus, Thomas Gregor, and William Bialek. Positional information, in bits. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41), 2013.

[59] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–1102, November 2010.

[60] Gerardo Aquino, Luke Tweedy, Doris Heinrich, and Robert G. Endres. Memory improves precision of cell sensing in fluctuating environments. *Scientific Reports*, 4:5688, July 2014.

[61] Massimo Vergassola, Emmanuel Villermaux, and Boris I Shraiman. Infotaxis as a strategy for searching without gradients. *Nature*, 445(7126):406–9, January 2007. ISSN 1476-4687.

[62] Eric D Siggia and Massimo Vergassola. Decisions on the fly in cellular sensory systems. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39):E3704–12, September 2013. ISSN 1091-6490.

[63] Georg Rieckh and Gašper Tkačik. Noise and information transmission in promoters with multiple internal states. *Biophysical Journal*, 106:1194–1204, 2014.

[64] Thomas R. Sokolowski and Gašper Tkačik. Optimizing information flow in small genetic networks. iv. spatial coupling. http://arxiv.org/abs/1501.04015, 2015.

[65] Jeremy E. Niven and Simon B. Laughlin. Energy limitation as a selective pressure on the evolution of sensory systems. *The Journal of Experimental Biology*, 211: 1792–1804, 2008.

[66] K. Kalyanasundaram and M. Graetzel. Artificial photosynthesis: biomimetic approaches to solar energy conversion and storage. *Current Opinion in Biotechnology*, 21:298–310, 2010.

[67] Sung Kuk Lee, Howard Chou, Timothy S. Ham, Taek Soon Lee, and Jay D Keasling. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology*, 19(6):556–563, 2008.

[68] Gavin E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90:1481–1487, 1998.

[69] Yuhai Tu. The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by maxwell's demons. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33):11737–11741, 2008.

[70] Ganhui Lan, Pablo Sartori, Silke Neumann, Victor Sourjik, and Yuhai Tu. The energy-speed-accuracy trade-off in sensory adaptation. *Nature Physics*, 8:422–428, March 2012.

[71] Pankaj Mehta and David J. Schwabb. Energetic costs of cellular computation. *Proceedings of the National Academy of Sciences of the United States of America*, 109:17978–17982, 2012.

[72] A. C. Barato, D. Hartich, and U. Seifert. Information-theoretic versus thermodynamic entropy production in autonomous sensory networks. *Physical Review E*, 87: 042104, 2013.

[73] Udo Seifert Andre C. Barato, David Hartich. Efficiency of cellular information processing. *New Journal of Physics*, 16:103024, October 2014.

[74] Stefano Bo, Marco Del Giudice, and A. Celani. Thermodynamic limits to information harvesting by sensory systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2015:P01014, 2015.

[75] Christopher C. Govern and Pieter Rein ten Wolde. Energy dissipation and noise correlations in biochemical sensing. *Physical Review Letters*, 113:258102, 2014.

[76] Pablo Sartori, Léo Granger, Chiu Fan Lee, and Jordan M. Horowitz. Thermodynamic costs of information processing in sensory adaptation. *PLOS Computational Biology*, 10(2):e1003974, December 2014.

[77] Nils B. Becker, Andrew Mugler, and Pieter Rein ten Wolde. Prediction and dissipation in biochemical sensing. http://arxiv.org/abs/1312.5625.

[78] Susanne Still, David A. Sivak, Anthony J. Bell, and Gavin E. Crooks. Thermodynamics of prediction. *Physical Review Letters*, 109:120604, September 2012.

[79] Linda A. Egger, Heiyoung Park, and Masayori Inouye. Signal transduction via the histidyl-aspartyl phosphorelay. *Genes to Cells*, 2:167–184, 1997.

[80] Jeffry B. Stock, Michael G. Surette, Mikhail Levit, and Peter Park. *Two-component Signal Transduction*, chapter 3, pages 25–51. ASM Press, 1995.

[81] Ann M. Stock, Victoria L. Robinson, and Paul N. Goudreau. Two-component signal transduction. *Annual Review of Biochemistry*, 69:183, 2000.

[82] Christopher M. Barbieri and Ann M. Stock. Comprehensive analysis of ompr phosphorylation, dimerization, and dna binding supports a canonical model for activation. *Journal of Molecular Biology*, 425:1612–1626, 2013.

[83] Arnim Weber, Stephanie A. Kögl, , and Kirsten Jung. Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in escherichia coli. *Journal of Bacteriology*, 188(20):7165–7175, 2006.

[84] John E. Lisman. A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 82:3055–3057, 1985.

[85] Wen Xiong and James E. Ferrell Jr. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426:460–465, 2003.

[86] Keiko Tanaka and George J. Augustine. A positive feedback signal transduction loop determines timing of cerebellar long-term depression. *Neuron*, 59:608–620, 2008.

[87] Olivier Cinquin and Jacques Demongeot. Roles of positive and negative feedback in biological systems. *Comptes Rendus Biologies*, 325(11):1085–1095, 2002.

[88] Bor-Sen Chen, Chia-Hung Chang, and Yun-Jen Chuang. Robust model matching control of immune systems under environmental disturbances: Dynamic game approach. *Journal of Theoretical Biology*, 253:824–837, 2008.

[89] Bor-Sen Chen, Chia-Hung Chang, and Hsiao-Ching Lee. Robust synthetic biology design: stochastic game theory approach. *Bioinformatics*, 25(14):1822–1830, 2009.

[90] J. von Neumann and O. Morgenstern. *Theory of Game and Economic Behavior*. Princeton University Press, 1944.

[91] Christopher C. Govern and Pieter Rein ten Wolde. Optimal resource allocation in cellular sensing systems. *Proceedings of the National Academy of Sciences of the United States of America*, 111(49):17486–17491, 2014.