*Article*

# Information Thermodynamics for Time Series of Signal-Response Models

**Andrea Auconi** [1], **Andrea Giansanti** [2,3] and **Edda Klipp** [1,*]

[1] Theoretische Biophysik, Humboldt-Universität zu Berlin, Invalidenstraße 42, D-10115 Berlin, Germany; andrea.auconi@gmail.com
[2] Dipartimento di Fisica, Sapienza Università di Roma, 00185 Rome, Italy; andrea.giansanti@roma1.infn.it
[3] INFN, Sezione di Roma 1, 00185 Rome, Italy
[*] Correspondence: edda.klipp@rz.hu-berlin.de

check for updates

**Abstract:** The entropy production in stochastic dynamical systems is linked to the structure of their causal representation in terms of Bayesian networks. Such a connection was formalized for bipartite (or multipartite) systems with an integral fluctuation theorem in [Phys. Rev. Lett. 111, 180603 (2013)]. Here we introduce the information thermodynamics for time series, that are non-bipartite in general, and we show that the link between irreversibility and information can only result from an incomplete causal representation. In particular, we consider a backward transfer entropy lower bound to the conditional time series irreversibility that is induced by the absence of feedback in signal-response models. We study such a relation in a linear signal-response model providing analytical solutions, and in a nonlinear biological model of receptor-ligand systems where the time series irreversibility measures the signaling efficiency.

**Keywords:** irreversibility; fluctuation theorems; time series; transfer entropy; causal influence

## 1. Introduction

The irreversibility of a process is the possibility to infer the existence of a time's arrow looking at an ensemble of realizations of its dynamics [1–3]. This concept appears in the nonequilibrium thermodynamics quantification of dissipated work or entropy production [4–6], and it relates the probability of paths with their time-reversal conjugates [7].

Fluctuation theorems have been developed to describe the statistical properties of the entropy production and its relation to information-theoretic quantities in both Hamiltonian and Langevin dynamics [8–10]. Particular attention was given to measurement-feedback controlled models [11,12] inspired by the Maxwell's demon [13], a gedanken-experiment in which mechanical work is extracted from thermodynamic systems using information. An ongoing experimental effort is put in the design and optimization of such information engines [14–17].

Theoretical studies clarified the role of fluctuations in feedback processes described by bipartite (or multipartite) stochastic dynamics, where fluctuation theorems set lower bounds on the entropy production of subsystems in terms of the Horowitz-Esposito information flow [18–20], or in terms of the transfer entropy [21,22] in the interaction between subsystems. Those inequalities form the second law of information thermodynamics [23], whose latest generalization was given in the form of geometrical projections into local reversible dynamics manifolds [24,25].

Time series can be obtained measuring continuous underlying dynamics at a finite frequency $\frac{1}{\tau}$, and this is the case of most real data. A measure of irreversibility for time series was defined in [26] as the Kullback–Leibler divergence [27] between the probability density of a time series realization and that of its time-reversal conjugate. Time series are non-bipartite in general, and this prevents

an unambiguous identification and additive separation of the physical entropy production or heat exchanged with thermal baths [28]. Then the time series irreversibility does not generally converge to the physical bipartite (or multipartite) entropy production in the limit of high sampling frequency $\tau \to 0$, except for special cases such as Langevin systems with constant diffusion coefficients, as we will discuss here.

The time series irreversibility measure defined in [26] depends on the statistical properties of the whole time series for non-Markovian processes. We define a measure of irreversibility that considers only the statistics of single transitions, and we call it *mapping irreversibility*. It recovers the irreversibility of whole time series only for Markovian systems. We study fluctuations of the mapping irreversibility introducing its stochastic counterpart. In the bivariate case of two interacting variables $x$ and $y$, we study the conditional stochastic mapping irreversibility defined as the difference between that of the joint process and that of a single marginal subsystem [29].

We define signal-response models as continuous-time stationary stochastic processes characterized by the absence of feedback. In the bidimensional case, a signal-response model consists of a fluctuating signal $x$ and a dynamic response $y$. In a recent work [30] we studied the information processing properties of linear (multidimensional) signal-response models. In that framework we defined a measure of causal influence to quantify how the macroscopic effects of asymmetric interactions are observed over time.

The backward transfer entropy is the standard transfer entropy [27] calculated in the ensemble of time-reversed trajectories. It was already shown to have a role in stochastic thermodynamics, in gambling theory, and in anti-causal linear regression models [21].

We derive an integral fluctuation theorem for time series of signal-response models that involves the backward transfer entropy. From this follows the II Law of Information thermodynamics for signal-response models, i.e., that the backward transfer entropy of the response $y$ on the past of the signal $x$ is a lower bound to the conditional mapping irreversibility. Please note that in general the conditional entropy production is zero lower bounded. Then our result shows that in time series it is the asymmetry of the interaction between subsystems $x$ and $y$, namely the absence of feedback that links irreversibility with information.

For the basic linear signal-response model (BLRM) discussed in [30], in the limit of small observational time $\tau$ the backward transfer entropy converges to the causal influence. Also, in the BLRM, we find that the causal influence rate converges to the Horowitz-Esposito [18] information flow.

A key quantity here is the observational time $\tau$, and the detection of the irreversibility of processes from real (finite) time series data is based on a fine-tuning of this parameter. We introduce such discussion with a biological model of receptor-ligand systems, where the entropy production measures the robustness of signaling.

The motivation of the present work is a future application of the stochastic thermodynamics framework to the analysis of time series data in biology and finance.

The paper is structured as follows. In Section 2 we provide a general introduction to stochastic thermodynamics introducing the irreversible entropy production. In Section 3 we state the setting and formalism we use for the treatment of bivariate time series, we define the stochastic (conditional) mapping irreversibility, we introduce the irreversibility density, and we review the general integral fluctuation theorem [31]. In Section 4 we discuss the integral fluctuation theorem for signal-response models involving the backward transfer entropy, and in Sections 5.1 and 5.2 we show how it applies to the BLRM, and to a biological model of receptor-ligand systems. In the Discussion section we review the results and we motivate further our work. We provide an Appendix section where the analytical results for the mapping irreversibility, for the backward transfer entropy, and for the Horowitz-Esposito information flow in the BLRM are discussed.

## 2. Introduction to Continuous Information Thermodynamics

### 2.1. Entropy Production in Heat Baths

Let us consider an ensemble of trajectories generated by a Markovian (memoryless) continuous-time stochastic process composed of two interacting variables $x$ and $y$ subject to Brownian noise $dW$. The stochastic differential equations (SDEs) describing such kind of processes can be written in the Ito interpretation [32] as:

$$
\begin{cases}
dx = g_x(x,y)dt + \sqrt{D_x(x,y)}\; dW_x \\
dy = g_y(x,y)dt + \sqrt{D_y(x,y)}\; dW_y
\end{cases}
\tag{1}
$$

where $D_x(x,y)$ and $D_y(x,y)$ are diffusion coefficients whose $(x,y)$ dependence takes into account the case of multiplicative noise. Brownian motion is characterized by $\langle dW_i(t)dW_j(t')\rangle = \delta_{ij}\delta_{tt'}dt$. The dynamics in (1) is bipartite, which means conditionally independent in updating: $p(x_{t+dt}, y_{t+dt}|x_t, y_t) = p(x_{t+dt}|x_t, y_t) \cdot p(y_{t+dt}|x_t, y_t)$.

The bipartite structure of (1) is fundamental in stochastic thermodynamics, because it allows the identification [28] and additive separation [18] of the heat exchanged with thermal baths in separate contact with $x$ and $y$ subsystems, $ds_b = ds_b^x + ds_b^y$. These are given by the detailed balance relation [5,10]:

$$
ds_b^x = \ln \frac{p(x_{t+dt}|x_t, y_t)}{p(\widetilde{x_{t+dt}}|\widetilde{x}_t, y_t)},
\tag{2}
$$

where $\widetilde{x_{t+dt}}$ is defined as the event of variable $x$ assuming value $x_t$ at time $t + dt$, and similarly $\widetilde{x}_t$ is the event of variable $x$ assuming value $x_{t+dt}$ at time $t$. An analogous expression to (2) holds for subsystem $y$. Time-integrals of the updating probabilities $p(x_{t+dt}|x_t, y_t)$ and $p(\widetilde{x_{t+dt}}|\widetilde{x}_t, y_t)$ can be written in terms of the SDE (1) using Onsager-Machlup action functionals [19,33].

Stochastic thermodynamics quantities are defined in single realizations of the probabilistic dynamics, in relation to the ensemble distribution [31,34]. As an example, the stochastic joint entropy is $s_{xy} = -\ln p_t(x_t, y_t)$, and its thermal (ensemble) average is the macroscopic entropy $S_{xy} = \langle s_{xy}\rangle_{p_t(x_t,y_t)}$. The explicit time dependence of $p_t$ describes the ensemble dynamics that in stationary processes is a relaxation to steady state. A SDE system such as (1) can be transformed into an equivalent partial differential equation in terms of probability currents [35], that is, the Fokker-Planck equation $\partial_t p(x,y,t) = -\partial_x J_x(x,y,t) - \partial_y J_y(x,y,t)$. Probability currents are related to average velocities with $J_x(x,y,t) \equiv p_t(x_t = x, y_t = y)\langle \dot{x}_t\rangle|_{x,y}$.

### 2.2. Feedbacks and Information

The stochastic entropy of subsystem $x$ *unaware* of the other subsystem $y$ is $s_x = -\ln p_t(x_t)$, and its time variation is $ds_x = \ln \frac{p_t(x_t)}{p_{t+dt}(x_{t+dt})}$. The apparent entropy production of subsystem $x$ with its heat bath is $ds_{x+b} = ds_x + ds_b^x$, and its thermal average $\langle ds_{x+b}\rangle$ can be negative due to the interaction with $y$, in apparent violation of the thermodynamics II Law. This is the case of Maxwell's demon strategies [11,12], where information on $x$ gained by the measuring device $y$ is exploited to exert a feedback force and extract work from $x$, such as in the feedback cooling of a Brownian particle [19,36]. Integral fluctuation theorems [10,23] (IFTs) provide lower bounds on the subsystems' macroscopic entropy production and extracted work in terms of information-theoretic measures [27]. The stochastic mutual information is defined as $I_t^{st} \equiv \ln\left(\frac{p_t(x_t,y_t)}{p_t(x_t)p_t(y_t)}\right)$, where "$st$" stands for stochastic. Its time derivative can be separated into contributions corresponding to single trajectory movements and ensemble probability currents in the two directions, $d_t I_t^{st} = i_t^x + i_t^y$. The stochastic information flux in the $x$ direction has the form $i_t^x = \dot{x}_t\partial_{x_t}\ln p_t(y_t|x_t) + \frac{\partial_x J_x(x,t)}{p_t(x_t)} - \frac{\partial_x J_x(x,y,t)}{p_t(x_t,y_t)}$, where $J_x(x,t) = \int dy\; J_x(x,y,t)$.

An analogous expression holds for $i_t^y$. The thermal average $I_{x \to y}(t) \equiv \langle i_t^y \rangle$ is the Horowitz-Esposito information flow [18,36]. At steady state it takes the form:

$$I_{x \to y} = \int \int dx dy \, J_y(x,y) \frac{\partial \ln p(x|y)}{\partial y}. \tag{3}$$

A recent formulation [19] upper bounds the average work extracted in feedback systems that in the steady-state bipartite framework is proportional to the $x$ bath entropy change $\langle ds_b^x \rangle$, with the information flow (3) towards the $y$ sensor. Such a result is recovered with a different formulation in terms of transfer entropies, and it is the Ito inequality [10,21,24] that reads:

$$-\langle ds_b^x \rangle \leq T_{x \to y}(dt) - T_{x \to y}(-dt) = dt I_{x \to y}. \tag{4}$$

where forward and backward stochastic transfer entropy [22] are respectively defined as $T_{x \to y}^{st}(dt) = \ln \left( \frac{p(y_{t+dt}|x_t,y_t)}{p(y_{t+dt}|y_t)} \right)$, and $T_{x \to y}^{st}(-dt) = \ln \left( \frac{p(y_t|x_{t+dt},y_{t+dt})}{p(y_t|y_{t+dt})} \right)$.

### 2.3. Irreversible Entropy Production

The stochastic (total) *irreversible* entropy production [18,24,36] of the joint system and thermal baths is:

$$ds_i^{xy} = ds_{xy} + ds_b^x + ds_b^y, \tag{5}$$

where $ds_{xy} = \ln \frac{p_t(x_t,y_t)}{p_{t+dt}(x_{t+dt},y_{t+dt})}$ is the joint system stochastic entropy change. If the ensemble is at steady state $p_{t+dt}(x_{t+dt}, y_{t+dt}) = p_t(x_{t+dt}, y_{t+dt}) = p(\tilde{x}_t, \tilde{y}_t)$. If we further assume that diffusion coefficients in (1) are nonzero constants, and this is the case of Langevin systems [28] where these are proportional to the temperature, then the conditional probability $p(\widetilde{x_{t+dt}}|\tilde{x}_t, y_t)$ is equivalent to $p(\widetilde{x_{t+dt}}|\tilde{x}_t, y(t) = y_{t+dt}) = p(\widetilde{x_{t+dt}}|\tilde{x}_t, \tilde{y}_t)$ under the time integral sign [10]. More precisely the term $\frac{1}{dt} \ln \left( \frac{p(\widetilde{x_{t+dt}}|\tilde{x}_t, y_t)}{p(\widetilde{x_{t+dt}}|\tilde{x}_t, \tilde{y}_t)} \right)$ almost surely vanishes. Then the irreversible entropy production (5) takes the form:

$$ds_i^{xy} = \ln \left( \frac{p(x_t, y_t, x_{t+dt}, y_{t+dt})}{p(\tilde{x}_t, \tilde{y}_t, \widetilde{x_{t+dt}}, \widetilde{y_{t+dt}})} \right). \tag{6}$$

Equation (6) shows the connection between entropy production and irreversibility of trajectories. The thermal average has the form of a Kullback–Leibler divergence [26,27] and satisfies $dS_i^{xy} \equiv \langle ds_i^{xy} \rangle \geq 0$. Using the Ito inequality [10,21,24] for both $ds_b^x$ and $ds_b^y$ does not lead to a positive lower bound in continuous (bipartite) stationary processes: $dS_i^{xy} \geq T_{x \to y}(-dt) - T_{x \to y}(dt) + T_{y \to x}(-dt) - T_{y \to x}(dt) = 0$. Nevertheless, it is clear that the irreversible entropy production $dS_i^{xy}$ is strictly positive when the interaction between subsystems is nonconservative [9]. Our main interest is the stationary dissipation due to asymmetric interactions between subsystems, and how it is manifested in time series.

## 3. Bivariate Time Series Information Thermodynamics

### 3.1. Setting and Definition of Causal Representations

Let us assume that we can measure the state of the system $(x,y)$ at a frequency $\frac{1}{\tau}$, thus obtaining time series. The finite observational time $\tau > 0$ makes the updating probability *not* bipartite: $p(x_{t+\tau}, y_{t+\tau}|x_t, y_t) = p(x_{t+\tau}|x_t, y_t, y_{t+\tau}) \cdot p(y_{t+\tau}|x_t, y_t) = p(x_{t+\tau}|x_t, y_t) \cdot p(y_{t+\tau}|x_t, y_t, x_{t+\tau})$. Therefore, a clear identification of thermodynamics quantities in time series is not possible. Let us take the Markovian SDE system (1) as the underlying process, and let us further assume stationarity. Then the statistical properties of time series obtained from a time discretization can be represented in the form of Bayesian networks, where links correspond to the way in which the joint probability density $p(x_t, y_t, x_{t+\tau}, y_{t+\tau})$ of states at the two instants $t$ and $t + \tau$ is factorized. Still, there are multiple ways of such factorization. We say that a Bayesian network is a *causal representation* of the dynamics if

conditional probabilities are expressed in a way that variables at time $t + \tau$ depend on variables at the same time instant or on variables at the previous time instant $t$ (and not vice-versa), and that the dependence structure is done in order to minimize the total number of conditions on the probabilities. This corresponds to a minimization of the number of links in the Bayesian network describing the dynamics with observational time $\tau$. Importantly, the causal representation is a way of factorizing the joint probability $p(x_t, y_t, x_{t+\tau}, y_{t+\tau})$, and not a claim of causality between observables.

We define the combination $\zeta_\tau^{xy}$ as a pair of successive states of the joint system $(x, y)$ separated by a time interval $\tau$, $\zeta_\tau^{xy} \equiv (x(t) = x_t, y(t) = y_t, x(t+\tau) = x_{t+\tau}, y(t+\tau) = y_{t+\tau}) \equiv f_\tau^{xy}(x_t, y_t, x_{t+\tau}, y_{t+\tau}) \equiv (x_t, y_t, x_{t+\tau}, y_{t+\tau})$. We use the identity functional $f_\tau^{xy}(a, b, c, d) \equiv (x(t) = a, y(t) = b, x(t+\tau) = c, y(t+\tau) = d)$ for an unambiguous specification of the backward combination $\widetilde{\zeta_\tau^{xy}}$. This is defined as the time-reversed conjugate of the combination $\zeta_\tau^{xy}$, meaning the inverted pair of the same two successive states, $\widetilde{\zeta_\tau^{xy}} \equiv f_\tau^{xy}(x_{t+\tau}, y_{t+\tau}, x_t, y_t) \equiv (\widetilde{x}_t, \widetilde{y}_t, \widetilde{x_{t+\tau}}, \widetilde{y_{t+\tau}})$. We defined backward variables of the type $\widetilde{x}_t$ meaning $x(t) = x_{t+\tau}$, such correspondences being possible only when states at both times $t$ and $t + \tau$ are given. The subsystems variables and backward variables are similarly defined as $\zeta_\tau^x = (x_t, x_{t+\tau})$, $\zeta_\tau^y = (y_t, y_{t+\tau})$, $\widetilde{\zeta_\tau^x} = (\widetilde{x}_t, \widetilde{x_{t+\tau}})$, and $\widetilde{\zeta_\tau^y} = (\widetilde{y}_t, \widetilde{y_{t+\tau}})$.

### 3.2. Definition of Mapping Irreversibility and the Standard Integral Fluctuation Theorem

A measure of coarse-grained entropy production for time series can be defined replacing $dt$ with the nonzero observational time $\tau$ in the general expression (5) obtaining:

$$\Delta s_i^{xy} = \Delta s_{xy} + \Delta s_b^x + \Delta s_b^y \equiv$$
$$\equiv \ln \frac{p(x_t, y_t)}{p(x_{t+\tau}, y_{t+\tau})} + \ln \frac{p(x_{t+\tau}|x_t, y_t)}{p(\widetilde{x_{t+\tau}}|\widetilde{x}_t, y_t)} + \ln \frac{p(y_{t+\tau}|y_t, x_t)}{p(\widetilde{y_{t+\tau}}|\widetilde{y}_t, x_t)}, \tag{7}$$

where we assumed stationarity, $p_t = p$. By definition $\Delta s_i^{xy}$ converges to the physical entropy production in the limit $\tau \to 0$, and it is a lower bound to it [37]. Importantly, such coarse-grained entropy production cannot have the form of an irreversibility measure such as (6) because $p(y_{t+\tau}|y_t, x_t, x_{t+\tau}) \neq p(y_{t+\tau}|y_t, x_t)$. With "irreversibility form" we mean that its thermal average is a Kullback–Leibler divergence measuring the distinguishability between forward and time-reverse paths. Therefore, we decided to keep the widely accepted time series irreversibility definition given in [26], in its form for stationary Markovian systems. Anyway, we are interested in the time-reversal asymmetry of time series from even more general models or data where no identification of thermodynamic quantities is required.

For the study of fluctuations, we define the stochastic mapping irreversibility with observational time $\tau$ of the joint system $(x, y)$ as:

$$\varphi_\tau^{xy} = \ln \left( \frac{p(\zeta_\tau^{xy})}{p(\widetilde{\zeta_\tau^{xy}})} \right). \tag{8}$$

The thermal average is called mapping irreversibility, $\Phi_\tau^{xy} \equiv \left\langle \varphi_\tau^{xy} \right\rangle_{p(\zeta_\tau^{xy})}$, and it describes the statistical properties of a single transition over an interval $\tau$. However, since the underlying dynamics (1) is Markovian, it describes the irreversibility of arbitrary long time series.

Let us note that $\varphi_\tau^{xy}$ does not generally converge to the total physical entropy production (5) in the limit $\tau \to 0$ because of the non-bipartite structure of $p(x_{t+\tau}, y_{t+\tau}|x_t, y_t)$ for any $\tau > 0$. It does converge anyway in most physical situations where bipartite underlying dynamics such as (1) has constant and strictly positive diffusion coefficients, and this is the case of Langevin systems. This is because the Brownian increments $dW_x$ and $dW_y$ are dominating the dynamics for small intervals $\tau$, then the estimate of $g_y(x_{t+t'}, y_{t+t'})$ (with $0 \leq t' \leq \tau$) based on $(x_t, y_t)$ is improved with the knowledge of $x_{t+\tau}$ just by a term of order $\partial_x g_y(x, y) \cdot W_x(\tau) \sim \sqrt{\tau}$, where we assumed a smooth $g_y(x, y)$. Therefore in

the limit $\tau \to 0$ it is almost surely $p(y_{t+\tau}|x_t, y_t, x_{t+\tau}) \to p(y_{t+\tau}|x_t, y_t)$ and $\varphi_\tau^{xy} \to \Delta s_i^{xy} \to ds_i^{xy}$, see Appendix D.

The stochastic mapping irreversibility satisfies the standard integral fluctuation theorem [31], i.e.,:

$$\left\langle e^{-\varphi_\tau^{xy}} \right\rangle_{p(\zeta_\tau^{xy})} = \int_\Omega d\zeta_\tau^{xy} p(\widetilde{\zeta_\tau^{xy}}) = 1, \tag{9}$$

where $d\zeta_\tau^{xy} = dx_t dy_t dx_{t+\tau} dy_{t+\tau}$, $\widetilde{dx_t} = dx_{t+\tau}$, and $\Omega$ is the whole space of the combination $\zeta_\tau^{xy}$. From the convexity of the exponential function it follows that the mapping irreversibility $\Phi_\tau^{xy}$ is non-negative. This is the standard thermodynamics II Law inequality for the joint system $(x, y)$ time series:

$$\Phi_\tau^{xy} = \left\langle \varphi_\tau^{xy} \right\rangle_{p(\zeta_\tau^{xy})} \geq 0. \tag{10}$$

Similarly, we define the stochastic mapping irreversibility for the two subsystems as $\varphi_\tau^x \equiv \ln\left(\frac{p(\zeta_\tau^x)}{p(\widetilde{\zeta_\tau^x})}\right)$ and $\varphi_\tau^y \equiv \ln\left(\frac{p(\zeta_\tau^y)}{p(\widetilde{\zeta_\tau^y})}\right)$, these being called the marginals [29]. Their ensemble averages are respectively denoted $\Phi_\tau^x \geq 0$ and $\Phi_\tau^y \geq 0$, and they also satisfy the standard II Law.

Although bivariate time series derived from the joint process (1) are Markovian, the one-dimensional subsystems time series are generally not. This is because subsystems trajectories are a coarse-grained representation of the full dynamics, and to reproduce the statistical properties of those trajectories a non-Markovian dynamics has to be assumed. Therefore $\Phi_\tau^x$ and $\Phi_\tau^y$ are generally different from the irreversibility calculated on a whole time series. The mapping irreversibility $\Phi_\tau^x$ describes the statistical properties of the whole time series only if it is Markovian. This is surely the case if $x$ is not influenced by $y$ in (1), $\partial_y g_x(x, y) = \partial_y D_x(x, y) = 0$, and motivated our study of signal-response models [30].

We define the conditional mapping irreversibility of $y$ given $x$ as the difference between the mapping irreversibility of the joint system $(x, y)$ and the mapping irreversibility of system $x$ alone:

$$\Phi_\tau^{y|x} \equiv \Phi_\tau^{xy} - \Phi_\tau^x = \left\langle \ln\left(\frac{p(y_t, y_{t+\tau}|x_t, x_{t+\tau})}{p(\widetilde{y_t}, \widetilde{y_{t+\tau}}|\widetilde{x_t}, \widetilde{x_{t+\tau}})}\right) \right\rangle_{p(\zeta_\tau^{xy})}. \tag{11}$$

This can be considered as a time series generalization to the conditional entropy production introduced in [29]. Also, the conditional mapping irreversibility satisfies the standard integral fluctuation theorem and the corresponding II Law-like inequality $\Phi_\tau^{y|x} = \left\langle \varphi_\tau^{y|x} \right\rangle_{p(\zeta_\tau^{xy})} \geq 0$.

In the general case where the evolution of each variable is influenced by the other variable (Equation (1)), we have a complete causal representation resulting from the dynamics (Figure 1), meaning that all edges are present in the Bayesian network. Please note that the horizontal arrows are non-directed because of the factorization degeneracy, meaning that the causal representation is given by multiple network topologies. In the complete case (Figure 1) we were not able to provide a more accurate characterization of the conditional mapping irreversibility $\Phi_\tau^{y|x}$ than the one given by the standard fluctuation theorem and the corresponding II Law-like inequality, $\Phi_\tau^{y|x} \geq 0$.

Let us recall that the inequalities for continuous bipartite systems [10,18,19,21,24] apply to the apparent entropy production $\Delta s_{x+b} = \Delta s_x + \Delta s_b^x$, and do not influence the total irreversible entropy production $\Delta s_i^{xy}$. Similarly, those results do not influence the mapping irreversibility $\Phi_\tau^{xy}$ and the conditional mapping irreversibility $\Phi_\tau^{y|x}$, for which in general only the standard zero lower bound can be provided.

$$y_{t+\tau} \text{———} x_{t+\tau}$$
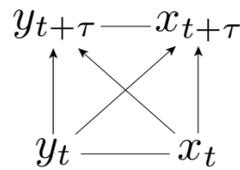
$$y_t \text{———} x_t$$

**Figure 1.** Complete causal representation. The arrows represent the way we factorize the joint probability density. In the complete case the causal representation is fully degenerate: $p(\zeta_\tau^{xy}) = p(x_t, y_t) \cdot p(x_{t+\tau}, y_{t+\tau}|x_t, y_t) = p(x_t, y_t) \cdot p(x_{t+\tau}|x_t, y_t) \cdot p(y_{t+\tau}|x_t, y_t, x_{t+\tau}) = p(x_t, y_t) \cdot p(y_{t+\tau}|x_t, y_t) \cdot p(x_{t+\tau}|x_t, y_t, y_{t+\tau})$.

We aim to relate the irreversibility of time series (8) to the (discrete) information flow between subsystems variables over time. We argue that fluctuation theorems linking the irreversibility of time series with information arise because of missing edges in the causal representation of the dynamics in terms of Bayesian networks. In the bivariate case there is only one class of time series generated from continuous underlying dynamics for which integral fluctuation theorems involving information measures can be written, and it corresponds to dynamics without feedback: the signal-response models.

*3.3. Ito Inequality for Time Series*

The Ito fluctuation theorem [10,21,24] for bipartite non-Markovian dynamics can be extended to Markovian non-bipartite time series if we modify the subsystems apparent entropy production $\Delta s_{y+b} = \Delta s_y + \Delta s_b^y$ into the explicitly non-bipartite form $\eta_\tau^{y|x}$:

$$\eta_\tau^{y|x} = \ln\left(\frac{p(y_{t+\tau}|y_t, x_t, x_{t+\tau})}{p(\widetilde{y_{t+\tau}}|\tilde{y}_t, \tilde{x}_t, \widetilde{x_{t+\tau}})}\right) + \ln\left(\frac{p(y_t)}{p(y_{t+\tau})}\right). \tag{12}$$

Then the Ito fluctuation theorem for time series is written:

$$\left\langle e^{-\eta_\tau^{y|x} + T_{y\to x}^{st}(-\tau) - T_{y\to x}^{st}(\tau) + I_{xy}^{st}(t+\tau) - I_{xy}^{st}(t)} \right\rangle_{p(\zeta_\tau^{xy})} = 1. \tag{13}$$

In stationary processes the mutual information is time invariant, $I_{xy}(t) = I_{xy}(t+\tau)$, and (13) implies the II Law-like inequality:

$$\langle \eta_\tau^{y|x} \rangle \geq T_{y\to x}(-\tau) - T_{y\to x}(\tau). \tag{14}$$

Similar to the apparent entropy production $\Delta s_y + \Delta s_b^y$ for bipartite systems, also $\langle \eta_\tau^{y|x} \rangle$ is not ensured to be positive if system $x$ acts like a Maxwell's demon.

The definition (12) does not have a clear physical meaning in time series, apart from its convergence to the apparent entropy production for $\tau \to 0$, again for a continuous underlying dynamics with constant nonzero diffusion coefficients. In addition, $\langle \eta_\tau^{y|x} \rangle$ does not have the form of a Kullback–Leibler divergence, and is then not considered a measure of irreversibility. Therefore, in the following we will be interested instead in the conditional mapping irreversibility $\Phi_\tau^{y|x}$ (Equation (11)). Importantly, there is no general connection between $\Phi_\tau^{y|x}$ and information measures, and such connection will result instead from the topology of the causal representation in signal-response models.

*3.4. The Mapping Irreversibility Density*

Let us use an equivalent representation of the mapping irreversibility in terms of backward probabilities [38] defined as $p_B(\zeta_\tau^{xy}) = p(x(t) = x_t, y(t) = y_t, x(t-\tau) = x_{t+\tau}, y(t-\tau) = y_{t+\tau})$. For

stationary processes it holds $p_B(\zeta_\tau^{xy}) = p(\widetilde{\zeta_\tau^{xy}})$ and $\varphi_\tau^{xy} = \ln(\frac{p(\zeta_\tau^{xy})}{p_B(\zeta_\tau^{xy})})$. We introduce here the mapping irreversibility density (with observational time $\tau$) for stationary processes as:

$$\psi(x_t, y_t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_{t+\tau}dy_{t+\tau}\, p(\zeta_\tau^{xy})\varphi_\tau^{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_{t+\tau}dy_{t+\tau}\, p(\zeta_\tau^{xy}) \ln\left(\frac{p(\zeta_\tau^{xy})}{p_B(\zeta_\tau^{xy})}\right) =$$

$$= p(x_t, y_t) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_{t+\tau}dy_{t+\tau}\, p(x_{t+\tau}, y_{t+\tau}|x_t, y_t) \ln\left(\frac{p(x(t+\tau)=x_{t+\tau},y(t+\tau)=y_{t+\tau}|x(t)=x_t,y(t)=y_t)}{p(x(t-\tau)=x_{t+\tau},y(t-\tau)=y_{t+\tau}|x(t)=x_t,y(t)=y_t)}\right). \quad (15)$$

The mapping irreversibility density $\psi(x_t, y_t)$ tells us which situations $(x_t, y_t)$ contribute more to the time series irreversibility of the macroscopic process. $\psi(x_t, y_t)$ is proportional to the distance (precisely to the Kullback–Leibler divergence [27]) of the distribution of future states $p(x_{t+\tau}, y_{t+\tau}|x_t, y_t)$ to the distribution of past states $p(x_{t-\tau}, y_{t-\tau}|x_t, y_t)$ of the same condition $(x_t, y_t)$.

## 4. The Fluctuation Theorem for Time Series of Signal-Response Models

If the system $(x, y)$ is such that the variable $y$ does not influence the dynamics of the variable $x$, then we are dealing with signal-response models (Figure 2). The stochastic differential equation for signal-response models is written in the Ito representation [32] as:

$$\begin{cases} dx = g_x(x)dt + \sqrt{D_x(x)}\ dW_x \\ dy = g_y(x,y)dt + \sqrt{D_y(x,y)}\ dW_y \end{cases} \quad (16)$$

The absence of feedback is written in $\frac{\partial g_x}{\partial y} = \frac{\partial D_x}{\partial y} = 0$. As a consequence, the conditional probability satisfies $p(y_t|x_t, x_{t+\tau}) = p(y_t|x_t)$, and the corresponding causal representation is incomplete, see the Bayesian network in Figure 2. In other words, signal-response models are specified by the property that $x_{t+\tau}$ is conditionally independent on $y_t$ given $x_t$.
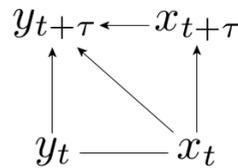


**Figure 2.** Causal representation of signal-response models. The joint probability density is factorized into $p(\zeta_\tau^{xy}) = p(x_t, y_t) \cdot p(x_{t+\tau}|x_t) \cdot p(y_{t+\tau}|x_t, y_t, x_{t+\tau})$.

For signal-response models we can provide a lower bound on the entropy production that is more informative than Equation (10), and that involves the backward transfer entropy $T_{y \to x}(-\tau)$. The backward transfer entropy [21] is a measure of discrete information flow towards the past, and is here defined as the standard transfer entropy for the ensemble of time-reversed combinations $\zeta_\tau^{xy}$. The stochastic counterpart as a function of $\zeta_\tau^{xy} \setminus y_t$ is defined as:

$$T_{y \to x}^{st}(-\tau) = \ln\left(\frac{p(x_t|y_{t+\tau}, x_{t+\tau})}{p(x_t|x_{t+\tau})}\right), \quad (17)$$

where $st$ stands for stochastic.

Then by definition $T_{y \to x}(-\tau) = \left\langle T_{y \to x}^{st}(-\tau) \right\rangle_{p(\zeta_\tau^{xy} \setminus y_t)}$. We keep the same symbol $T_{y \to x}$ as the standard transfer entropy because in stationary processes the backward transfer entropy is the standard transfer entropy (calculated on forward trajectories) for negative shifts $-\tau$.

The fluctuation theorem for time series of signal-response models is written:

$$\left\langle e^{-\varphi_\tau^{xy}+\varphi_\tau^x+T_{y\to x}^{st}(-\tau)} \right\rangle_{p(\zeta_\tau^{xy})} =$$

$$= \int_\Omega d\zeta_\tau^{xy} \, p(\widetilde{y_{t+\tau}}|\widetilde{x_t},\widetilde{y_t},\widetilde{x_{t+\tau}}) p(x_t, x_{t+\tau}, y_{t+\tau}) = 1, \tag{18}$$

where we used the signal-response property of no feedback $p(\widetilde{y_t}|\widetilde{x_t},\widetilde{x_{t+\tau}}) = p(\widetilde{y_t}|\widetilde{x_t})$, the stationarity property $p(\widetilde{y_t}|\widetilde{x_t}) = p(y_{t+\tau}|x_{t+\tau})$, the correspondence $dy_t = \widetilde{dy_{t+\tau}}$, and the normalization property $\int_{-\infty}^{\infty} d\widetilde{y_{t+\tau}} \, p(\widetilde{y_{t+\tau}}|\widetilde{x_t},\widetilde{y_t},\widetilde{x_{t+\tau}}) = 1$.

From the convexity of the exponential it follows the II Law of Information thermodynamics for time series of signal-response models:

$$\Phi_\tau^{y|x} \equiv \Phi_\tau^{xy} - \Phi_\tau^x \geq T_{y\to x}(-\tau), \tag{19}$$

and this is the central relation we wish to study in the Applications section hereafter.

In the limit of $\tau \to 0$ and constant nonzero diffusion coefficients, $\Phi_\tau^{y|x}$ converges to the heat exchanged with the thermal bath attached to subsystem $y$, $\langle ds_m^y \rangle$. Therefore, the inequality (19) converges to the Ito inequality (14) in its form for Markovian signal-response bipartite systems [21,24] as expected. Indeed, in signal-response models $\Phi_\tau^{y|x} = \langle \eta_\tau^{y|x} \rangle$, $T_{y\to x}(\tau) = 0$, and (14) transforms into (19). Therefore (19) can be regarded as an extension to (non-bipartite) time series of the Ito inequality [10,21,24]. This shows that in time series it is the asymmetry of the interaction between subsystems $x$ and $y$, namely the absence of feedback, that links irreversibility with information.

Please note that $\Phi_\tau^x$ is equivalent to the original time series irreversibility [26] because the $x$ time series is Markovian in the absence of feedback.

In causal representations of correlated stationary processes, the factorization of $p(x_t, y_t)$ is unnecessary, and only the structure of the transition probability $p(x_{t+\tau}, y_{t+\tau}|x_t, y_t)$ has to be specified. Please note that the direction of the horizontal $x_t$-$y_t$ arrow is never specified (see Figures 1 and 2). In the complete (symmetric) case with feedback we also do not specify the direction of the horizontal $x_{t+\tau}$-$y_{t+\tau}$ arrow because of the full degeneracy (see Figure 1). The importance of the causal representation is seen in signal-response models (Figure 2) because we could have decomposed the transition probability as well into the non-causal decomposition $p(x_{t+\tau}, y_{t+\tau}|x_t, y_t) = p(y_{t+\tau}|x_t, y_t) \cdot p(x_{t+\tau}|x_t, y_t, y_{t+\tau})$, but this does not lead to the fluctuation theorem (18).

## 5. Applications

### 5.1. The Basic Linear Response Model

We study the II Law for signal-response models (Equation (19)) in the BLRM, whose information processing properties are already discussed in [30]. The BLRM is composed of a fluctuating signal $x$ described by the Ornstein-Uhlenbeck process [39,40], and a dynamic linear response $y$ to this signal:

$$\begin{cases} dx = -\frac{x}{t_{rel}} dt + \sqrt{D} \, dW \\ \frac{dy}{dt} = \alpha x - \beta y \end{cases} \tag{20}$$

The response $y$ is considered in the limit of weak coupling with the thermal bath $D_y \to 0$, while the signal is attached to the source of noise, $D_x = D > 0$.

This model allows analytical representations for the mapping irreversibility $\Phi_\tau^{xy}$ (calculated in Appendix A) and the backward transfer entropy $T_{y\to x}(-\tau)$ (calculated in Appendix B). We find that once the observational time $\tau$ is specified, $\Phi_\tau^{xy}$ and $T_{y\to x}(-\tau)$ are both functions of just the two parameters $t_{rel}$ and $\beta$, which describe respectively the time scale of the fluctuations of the signal and the time scale of the response to a deterministic input. In addition, if we choose to rescale the time

units by $t_{rel}$ to compare fluctuations of different timescales, we find that irreversibility measures are function of just the product $\beta t_{rel}$ that is then the only free parameter in the model.

Since the signal is a time-symmetric (reversible) process, $\Phi_\tau^x = 0$, the backward transfer entropy $T_{y \to x}(-\tau)$ is the lower bound on the total entropy production $\Phi_\tau^{xy}$ in the BLRM.

The plot in Figure 3 shows the mapping irreversibility $\Phi_\tau^{xy}$ and the backward transfer entropy $T_{y \to x}(-\tau)$ as a function of the observational time $\tau$. In the limit of small $\tau$, the entropy production diverges because of the deterministic nature of the response dynamics (the standard deviation on the determination of the velocity $\frac{dy}{dt}$ due to instantaneous movements of the signal vanishes as $\alpha \sqrt{D} \sqrt{dt} \to 0$). The backward transfer entropy $T_{y \to x}(-\tau)$ instead vanishes for $\tau \to 0$ because the Brownian motion has nonzero quadratic variation [32] and is the dominating term in the signal dynamics for small time intervals. In the limit of large observational time intervals $\tau \to \infty$ the entropy production is asymptotically double the backward transfer entropy that is its lower bound given by the II Law for signal-response models (Equation (19)), $\frac{\Phi_\tau^{xy}}{T_{y \to x}(-\tau)} \to 2$. Interestingly, this limit of 2 is valid for any choice of the parameters in the BLRM.
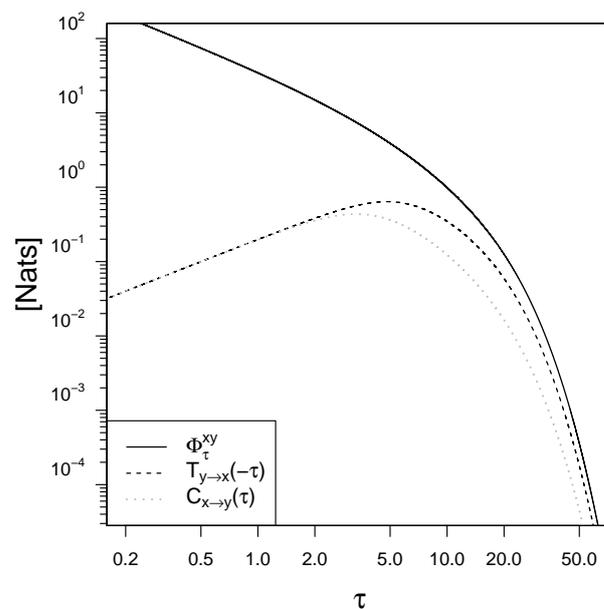


**Figure 3.** Mapping irreversibility $\Phi_\tau^{xy}$, backward transfer entropy $T_{y \to x}(-\tau)$ and causal influence $C_{x \to y}(\tau)$ in the BLRM as a function of the observational time interval $\tau$. The parameters are $\beta = 0.2$ and $t_{rel} = 10$. All graphs are produced using R [41].

Let us recall the definition of causal influence $C_{x \to y}(\tau)$ as a measure of information flow for time series [30]:

$$C_{x \to y}(\tau) \equiv I(x(t), y(t + \tau)) - R(\tau) \equiv$$
$$\equiv I(x(t), y(t + \tau)) - \frac{1}{2} \ln \left( \frac{e^{2(I(x_t, y_t) + I(y_{t+\tau}, (x_t, y_t)))}}{e^{2I(x_t, y_t)} + e^{2I(y_{t+\tau}, (x_t, y_t))} - 1} \right). \tag{21}$$

$R(\tau)$ is the redundancy measure quantifying that fraction of time-lagged mutual information $I(x_t, y_{t+\tau})$ that the signal $x_t$ gives on the evolution $y_{t+\tau}$ of the response that is already contained in the knowledge of the current state $y_t$ of the response [42].

Interestingly, for small observational time $\tau \to 0$, the causal influence of the signal on the evolution of the response converges to the backward transfer entropy of the response on the past of the signal $C_{x\to y}(\tau) \to T_{y\to x}(-\tau)$, and they both vanish with $\tau\beta$. Also, the causal influence rate defined as $\lim_{\tau\to 0} \frac{C_{x\to y}(\tau)}{\tau}$ converges to the Horowitz-Esposito [18] information flow $I^{x\to y}$ (details in Appendix C).

For large observational time $\tau \to \infty$ instead the causal influence converges to the standard (forward) transfer entropy $C_{x\to y}(\tau) \to T_{y\to x}(\tau)$. Also, in this limit $\tau \to \infty$, the causal influence is an eighth of the entropy production $\frac{\Phi_\tau^{xy}}{C_{x\to y}(\tau)} \to 8$ for any choice of the parameters in the BLRM.

In general, the limit $\tau \to \infty$ corresponds to recording the system state at a rate that is much slower compared to any stationary dynamics, so that only an exponentially small time-delayed information flow is observed. Similarly, time asymmetries in the dynamics become less visible and the irreversibility measures vanish.

Let us now consider the mapping irreversibility density $\psi(x_t, y_t)$ in the BLRM for small and large observational time $\tau$. In Figure 4 we choose a $\tau$ smaller than the characteristic response time $\frac{1}{\beta}$ and smaller than the characteristic time of fluctuations $t_{rel}$. In the limit $\tau \to 0$ the signal dynamics is dominated by noise and the entropy production is mainly given by movements of the response $y$. The two spots correspond to the points where the product of the density $p(x_t, y_t)$ times the absolute value of the instant velocity $\dot{y}$ is larger. For longer intervals $\tau \gtrsim \frac{1}{\beta}$ (that is the case of Figure 5) the history of the signal becomes relevant because it determined the present value of the response, therefore the irreversibility density is also distributed on those points of the diagonal (corresponding to roughly $\dot{y}_t = 0$) where the absolute value of the signal $x_t$ is big enough. Also, consequently, in this regime the backward transfer entropy is a meaningful lower bound to the entropy production, that is Equation (19).
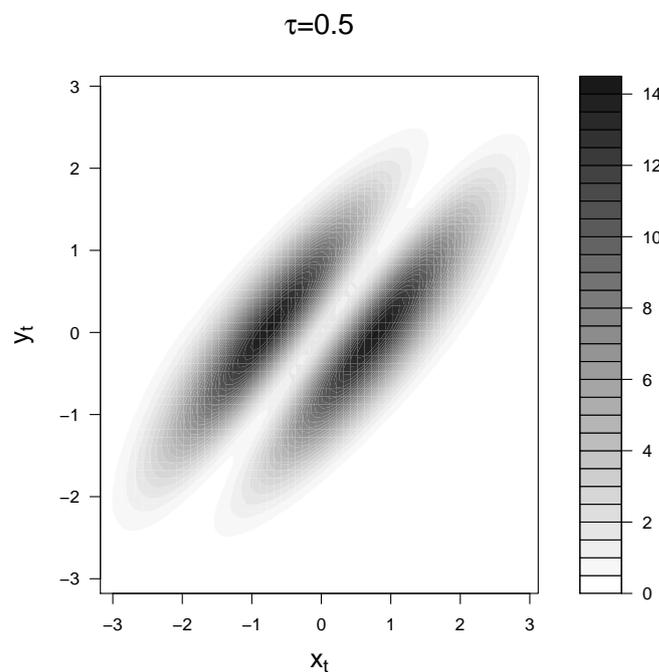


**Figure 4.** Mapping irreversibility density $\psi(x_t, y_t)$ for the BLRM at $\tau = 0.5 < \frac{1}{\beta} < t_{rel}$. The parameters are $\beta = 0.2$ and $t_{rel} = 10$. Both $\psi(x_t, y_t)$ and the space $(x, y)$ are expressed in units of standard deviations.
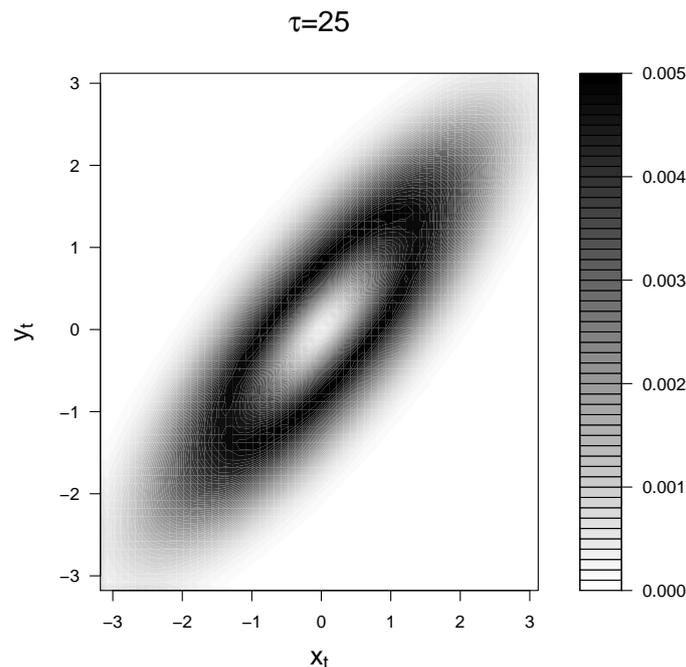
**Figure 5.** Mapping irreversibility density $\psi(x_t, y_t)$ for the BLRM at $\tau = 25 > t_{rel} > \frac{1}{\beta}$. The parameters are $\beta = 0.2$ and $t_{rel} = 10$. Both $\psi(x_t, y_t)$ and the space $(x, y)$ are expressed in units of standard deviations.

*5.2. Receptor-Ligand Systems*

The receptor-ligand interaction is the fundamental mechanism of molecular recognition in biology and is a recurring motif in signaling pathways [43,44]. The fraction of activated receptors is part of the cell's representation of the outside world, it is the cell's estimate on the concentration of ligands in the environment, upon which it bases its protein expression and response to external stress.

If we could experimentally keep the concentration of ligands fixed we would still get a fluctuating number of activated receptors due to the intrinsic stochasticity of the macroscopic description of chemical reactions. Recent studies allowed a theoretical understanding of the origins of the macroscopic "noise" (i.e., the output variance in the conditional probability distributions), and raised questions about the optimality of the input distributions in terms of information transmission [45–48].

Here we consider the dynamical aspects of information processing in receptor-ligand systems [49,50], where the response is integrated over time. If the perturbation of the receptor-ligand binding on the concentration of free ligands is negligible, that is in the limit of high ligand concentration, receptor-ligand systems can be modeled as nonlinear signal-response models [51]. We write our model of receptor-ligand systems in the Ito representation [32] as:

$$
\begin{cases}
dx = -(x-1)dt + x\, dW_x \\
dy = k_{on}(1-y)\frac{x^h}{1+x^h}dt - k_{off}y\,dt + y(1-y)dW_y
\end{cases}
\tag{22}
$$

The fluctuations of the ligand concentration $x$ are described by a mean-reverting geometric Brownian motion, with an average $\langle x \rangle = 1$ in arbitrary units. The response, that is the fraction of activated receptors $y$, is driven by a Hill-type interaction with the signal with cooperativity coefficient $h$, and chemical bound/unbound rates $k_{on}$ and $k_{off}$. For simplicity, the dynamic range of the response is set to be coincident with the mean value of the ligand concentration that means choosing a Hill constant

$K = \langle x \rangle = 1$. The form of the $y$ noise is set by the biological constraint $0 < y < 1$. For simplicity, we choose a cooperativity coefficient of $h = 2$ that is the smallest order of sigmoidal functions.

The mutual information between the concentration of ligands and the fraction of activated receptors in a cell is a natural choice for quantifying its sensory properties [52]. Here we argue that in the case of signal-response models, the conditional entropy production is the relevant measure, because it quantifies how the dynamics of the signal produces irreversible transitions in the dynamics of the response, which is closely related to the concept of causation. Besides, our measure of causal influence [30] has yet not been generalized to the nonlinear case, while the entropy production has a consistent thermodynamical interpretation [31].

We simulated the receptor-ligand model of Equation (22), and we evaluated numerically the mapping irreversibility $\Phi_\tau^{xy}$ and the backward transfer entropy $T_{y \to x}(-\tau)$ using a multivariate Gaussian approximation for the conditional probabilities $p(x_{t+\tau}, y_{t+\tau}|x_t, y_t)$ (details in Appendix E). The II Law for signal-response models sets $\Phi_\tau^{xy} \geq T_{y \to x}(-\tau)$ and proves to be a useful tool for receptor-ligand systems, as it is seen if Figure 6. Please note that the numerical estimation of the entropy production requires statistically many more samples compared to the backward transfer entropy: $\Phi_\tau^{xy}$ depends on $\zeta_\tau^{xy}$ (4 dimensions) while $T_{y \to x}(-\tau)$ depends on $\zeta_\tau^{xy} \setminus y_t$ (3 dimensions). In a real biological experimental setting, the sampling process is expensive, and the backward transfer entropy is therefore a useful lower bound for the entropy production, and an interesting characterization of the system to be used when the number of samples is not large enough.
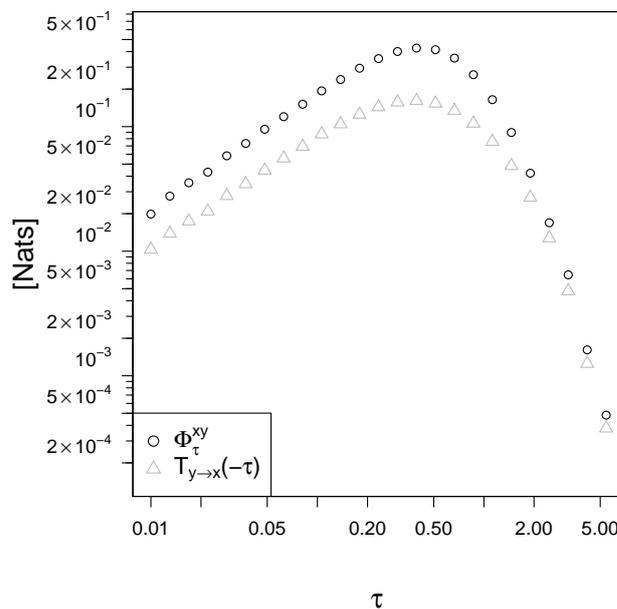


**Figure 6.** Mapping irreversibility and backward transfer entropy in our model of receptor-ligand systems (Equation (22)). The parameters are $k_{on} = 5$, $k_{off} = 1$, and $h = 2$.

The intrinsic noise of the response $y(1-y)dW_y$ is the dominant term in the response dynamics for small intervals $\tau$, and $xdW_x$ is the dominant term for the signal. This makes both $\Phi_\tau^{xy}$ and $T_{y \to x}(-\tau)$ vanish in the limit $\tau \to 0$. In the limit of large observational time $\tau$, as it is also the case for the BLRM and in any stationary process, the entropy production for the corresponding time series $\Phi_\tau^{xy}$ and all the information measures are vanishing, because the memory of the system is damped exponentially over time by the relaxation parameter $k_{off}$ ($\beta$ in the BLRM). Therefore, to better detect the irreversibility of a

process one must choose an appropriate observational time $\tau$. In the receptor-ligand model of Equation (22) with parameters $k_{on} = 5$, $k_{off} = 1$ and $h = 2$ we see that the optimal observational time is around $\tau \approx 0.5$ (see Figure 6). Here for "optimal" we mean the observational time that corresponds to the highest mapping irreversibility $\Phi_\tau^{xy}$. In general, one might be interested in inferring the irreversibility rate (that is $\frac{\Phi_\tau^{xy}}{\tau}$ in the limit $\tau \to 0$) looking at time series data with finite sampling interval $\tau$. In the receptor-ligand model of Figure 6 the irreversibility rate converges to 2.

## 6. Discussion

To put in perspective our work let us recall that the well-established integral fluctuation theorem for stochastic trajectories [34] leads to a total irreversible entropy production with zero lower bound, that is the standard II Law of thermodynamics. Modern inequalities such as the II Law of Information thermodynamics [21,23,24] describe how the information continuously shared between subsystems can lead to an "apparent" negative entropy production in (one of) the subsystems. Nevertheless, these do not bring to any difference in the total joint irreversible entropy production whose lower bound is still zero.

Our aim here was to characterize cases in which more informative lower bounds on the total irreversible entropy production can be provided. Ito-Sagawa [10] already showed that for Bayesian controlled systems (where a parameter can be varied to perform work) a general fluctuation theorem for the subsystems and the relative lower bound on entropy production is linked to the topology of the Bayesian network representation associated with the stochastic dynamics of the system. This connection seems to be even stronger in the case of the total joint (uncontrolled) system that is the object of our study. We show in the bidimensional case of a pair of signal-response variables how a missing arrow in the Bayesian network describing the dynamics leads to a fluctuation theorem.

The detailed fluctuation theorem linking work dissipation and the irreversibility of trajectories in nonequilibrium transformations [5,8] holds in mechanical systems attached to heat reservoirs. We are interested here in the irreversibility of trajectories in more general models, and especially those featuring asymmetric interactions, since that is a widespread feature in models of biological systems or in asset pricing models in quantitative finance. In particular, we do not adopt a Hamiltonian description of work and heat or a microscopic reversibility assumption, and the detailed fluctuation theorem (8) is, properly, not a theorem but itself a definition of irreversibility.

We study time series resulting from a discretization with observational time $\tau$ of continuous stochastic processes. Importantly the underlying bipartite process appears, at limited time resolution, as a non-bipartite process. As a consequence, there is no general convergence of the time series irreversibility to the physical entropy production except for special cases such as Langevin systems with nonzero constant diffusion coefficients. Our mapping irreversibility (8) is the Markovian approximation of the time series irreversibility definition given in [26]. While it is well defined for any stationary process, it describes the statistical properties of long time series only in the Markovian case.

For general interacting dynamics like (1) we are not able to provide a more significant lower bound to the mapping irreversibility than the standard II Law of thermodynamics (10). A more informative lower bound on the mapping irreversibility is found for signal-response models described by the absence of feedback. This sets the backward transfer entropy as a lower bound to the conditional entropy production, and describes the connection between the irreversibility of time series and the discrete information flow towards past between variables.

Importantly, the relation between irreversibility and information measures is not given in general for time series because the results on continuous bipartite systems do not generalize to the time series irreversibility. It appears exactly because of the absence of feedback, and of the corresponding non-complete causal representation. We restrict ourselves to the bivariate case here, but we conjecture that fluctuation theorems for multidimensional stochastic autonomous dynamics should arise in general as a consequence of missing arrows in the (non-complete, see e.g., Figure 2) causal representation of the dynamics in terms of Bayesian networks.

In our opinion, a general relation connecting the incompleteness of the causal representation of the dynamics and fluctuation theorems is still lacking.

Finally, let us note that exponential averages such as our integral fluctuation theorem (18) are dominated by (exponentially) rare realizations [53], and the corresponding II Law inequalities such as our (19) are often poorly saturated bounds. In the receptor-ligand model discussed in section II.B the backward transfer entropy lower bound is roughly one half of the mapping irreversibility, and this is also the case in the BLRM for large $\tau$ where the ratio converges exactly to $\frac{1}{2}$. This limitation is quite general, see for example the information thermodynamic bounds on signaling robustness given in [54].

We also introduced a discussion about the observational time $\tau$ in data analysis. In a biological model of receptor-ligand systems we showed that it must be fine-tuned for a robust detection of the irreversibility of the process, which is related to the concept of causation [30] and therefore to the efficiency of biological coupling between signaling and response.

**Author Contributions:** All authors contributed equally to this work.

**Abbreviations**

The following abbreviations are used in this manuscript:

OU       Ornstein-Uhlenbeck process
BLRM   basic linear signal-response model

**Appendix A. Mapping Irreversibility in the BLRM**

Let us consider an ensemble of stochastic trajectories generated with the BLRM (Equation (20)). The mapping irreversibility $\Phi_\tau^{xy}$ here is the Kullback–Leibler divergence [27] between the probability density $p(\zeta_\tau^{xy})$ of pairs of successive states $\zeta_\tau^{xy}$ separated by a time interval $\tau$ of the original (infinitely long and stationary) trajectory and the probability density $p_B(\zeta_\tau^{xy}) = p(\widetilde{\zeta_\tau^{xy}})$ of the same pairs of successive states $\zeta_\tau^{xy}$ of the time-reversed conjugate of the original trajectory (Equation (8)). For the sake of clarity, we use here in this appendix the full formalism rather than the compact one based on the functional form $f_\tau^{xy}$.

The time-reversed density of a particular couple of successive states, $(x(t) = \gamma, y(t) = \delta)$ and $(x(t + \tau) = \mu, y(t + \tau) = \xi)$, is equivalent to the original density of the exchanged couple of states, $(x(t) = \mu, y(t) = \xi)$ and $(x(t + \tau) = \gamma, y(t + \tau) = \delta)$. Therefore the density $p(\widetilde{\zeta_\tau^{xy}}) = p(x(t) = \mu, y(t) = \xi, x(t + \tau) = \gamma, y(t + \tau) = \delta)$ is the transpose of the density $p(\zeta_\tau^{xy}) = p(x(t) = \gamma, y(t) = \delta, x(t + \tau) = \mu, y(t + \tau) = \xi)$.

The mapping irreversibility for the BLRM is then written as:

$$\Phi_\tau^{xy} = \langle \varphi_\tau^{xy} \rangle_{p(\zeta_\tau^{xy})} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\gamma d\delta d\mu d\xi \, p\left(x(t) = \gamma, y(t) = \delta, x(t + \tau) = \mu, y(t + \tau) = \xi\right) *$$
$$* \ln\left(\frac{p(x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu, y(t+\tau) = \xi)}{p(x(t) = \mu, y(t) = \xi, x(t+\tau) = \gamma, y(t+\tau) = \delta)}\right). \tag{A1}$$

The BLRM is ergodic, therefore the densities $p(\zeta_\tau^{xy})$ and $p(\widetilde{\zeta_\tau^{xy}})$ can be empirically sampled looking at a single infinitely long trajectory.

The causal structure of the BLRM (and of any signal-response model, see Figure 2) is such that the evolution of the signal is not influenced by the response, $p(x(t + \tau)|x(t), y(t)) = p(x(t + \tau)|x(t))$.

Then we can write the joint probability densities $p(\zeta_\tau^{xy})$ of couples of successive states over a time interval $\tau$ of the original trajectory as:

$$p(\zeta_\tau^{xy}) \equiv p(x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu, y(t+\tau) = \xi) =$$
$$= p(x(t) = \gamma) \cdot p(y(t) = \delta | x(t) = \gamma) \cdot p(x(t+\tau) = \mu | x(t) = \gamma) *$$
$$* p(y(t+\tau) = \xi | x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu). \tag{A2}$$

We need to evaluate all these probabilities. Since we are dealing with linear models, these are all Gaussian distributed, and we will calculate only the expected value and the variance of the relevant variables involved.

The system is Markovian, $p(x(t+\tau)|x(t+t'), x(t)) = p(x(t+\tau)|x(t+t'))$ with $0 \leq t' \leq \tau$, and the Bayes rule assumes the form $p(x(t+t')|x(t), x(t+\tau)) = \frac{p(x(t+t')|x(t))p(x(t+\tau)|x(t+t'))}{p(x(t+\tau)|x(t))}$. Then we calculate the conditional expected value for the signal $x(t+t')$ given a condition for its past $x(t)$ and another condition for its future $x(t+\tau)$ as:

$$\langle x(t+t')|x(t), x(t+\tau) \rangle = x(t)e^{-\frac{t'}{t_{rel}}} \frac{1 - e^{-\frac{2(\tau-t')}{t_{rel}}}}{1 - e^{-\frac{2\tau}{t_{rel}}}} + x(t+\tau)e^{-\frac{\tau-t'}{t_{rel}}} \frac{1 - e^{-\frac{2t'}{t_{rel}}}}{1 - e^{-\frac{2\tau}{t_{rel}}}}. \tag{A3}$$

Now we can calculate the full-conditional expectation of the response:

$$\langle y(t+\tau)|x(t), y(t), x(t+\tau) \rangle = y(t)e^{-\beta\tau} + \alpha \int_0^\tau dt' e^{-\beta(\tau-t')} \langle x(t+t')|x(t), x(t+\tau) \rangle =$$
$$= y(t)e^{-\beta\tau} + \alpha \frac{e^{-\beta\tau}}{1-e^{-\frac{2\tau}{t_{rel}}}} \left( x(t)\left(\frac{e^{\tau(\beta-\frac{1}{t_{rel}})}-1}{\beta-\frac{1}{t_{rel}}} - \frac{e^{\tau(\beta-\frac{1}{t_{rel}})}-e^{-\frac{2\tau}{t_{rel}}}}{\beta+\frac{1}{t_{rel}}}\right) + x(t+\tau)\left(\frac{e^{\beta\tau}-e^{-\frac{\tau}{t_{rel}}}}{\beta+\frac{1}{t_{rel}}} - \frac{e^{\tau(\beta-\frac{2}{t_{rel}})}-e^{-\frac{\tau}{t_{rel}}}}{\beta-\frac{1}{t_{rel}}}\right) \right). \tag{A4}$$

One can immediately check that the limits for small and large time intervals $\tau$ verify respectively $\lim_{\tau\to 0} \langle y(t+\tau)|x(t), y(t), x(t+\tau) \rangle = y(t)$ and $\lim_{\tau\to\infty} \langle y(t+\tau)|x(t), y(t), x(t+\tau) \rangle = x(t+\tau)\frac{\alpha t_{rel}}{\beta t_{rel}+1} = \langle y(t+\tau)|x(t+\tau) \rangle$.

The causal order for the evolution of the signal is such that $p(x(t+t'')|x(t), x(t+t'), x(t+\tau)) = p(x(t+t'')|x(t+t'), x(t+\tau))$ if $0 \leq t' \leq t'' \leq \tau$. Then we can calculate:

$$\langle x(t+t')x(t+t'')|x(t), x(t+\tau) \rangle_{t'' \geq t'} =$$
$$= \int_{-\infty}^{\infty} dx(t+t') p(x(t+t')|x(t), x(t+\tau))x(t+t') \langle x(t+t'')|x(t+t'), x(t+\tau) \rangle =$$
$$= \langle x(t+t')|x(t), x(t+\tau) \rangle *$$
$$* \left( x(t+\tau)e^{-\frac{\tau-t''}{t_{rel}}} \frac{\sigma^2_{t''-t'}}{\sigma^2_{\tau-t'}} + e^{-\frac{t''-t'}{t_{rel}}} \frac{\sigma^2_{\tau-t''}}{\sigma^2_{\tau-t'}} \left( \frac{1}{x(t)\frac{e^{-\frac{t'}{t_{rel}}}}{\sigma^2_{t'}} + x(t+\tau)\frac{e^{-\frac{\tau-t'}{t_{rel}}}}{\sigma^2_{\tau-t'}}} + \frac{x(t)\frac{e^{-\frac{t'}{t_{rel}}}}{\sigma^2_{t'}} + x(t+\tau)\frac{e^{-\frac{\tau-t'}{t_{rel}}}}{\sigma^2_{\tau-t'}}}{\frac{1}{\sigma^2_{t'}} + \frac{e^{-\frac{2(\tau-t')}{t_{rel}}}}{\sigma^2_{\tau-t'}}} \right) \right). \tag{A5}$$

Let us write the full-conditional expectation of the squared response as a function of the expectations we just calculated:

$$\langle y^2(t+\tau)|x(t), y(t), x(t+\tau) \rangle = y^2(t)e^{-2\beta\tau} + 2\alpha y(t)e^{-2\beta\tau} \int_0^\tau dt' e^{\beta t'} \langle x(t+t')|x(t), x(t+\tau) \rangle +$$
$$+ \alpha^2 e^{-2\beta\tau} \int_0^\tau \int_0^\tau dt' dt'' e^{\beta(t'+t'')} \langle x(t+t')x(t+t'')|x(t), x(t+\tau) \rangle. \tag{A6}$$

A relevant feature of linear response models is that the conditional variances do not depend on the particular values of the conditioning variables [30]. Here we consider the full-conditional variance $\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)} = \langle y^2(t+\tau)|x(t), y(t), x(t+\tau) \rangle - \langle y(t+\tau)|x(t), y(t), x(t+\tau) \rangle^2$, and it will be

independent of the conditions $x(t)$, $y(t)$, and $x(t+\tau)$. Then the remaining terms in $\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)}$ sum up to:

$$\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)} = 2\frac{\alpha^2 e^{-2\beta\tau}}{\sigma_\tau^2}\int_0^\tau dt''\sigma^2_{\tau-t''}e^{t''(\beta-\frac{1}{t_{rel}})}\int_0^{t''}dt'\sigma^2_{t'}e^{t'(\beta+\frac{1}{t_{rel}})} = 2\alpha^2\sigma_x^2\frac{e^{-2\beta\tau}}{1-e^{-\frac{2\tau}{t_{rel}}}}*$$

$$*\left(-\frac{2}{t_{rel}}\frac{\beta+\frac{1}{t_{rel}}-\frac{2}{t_{rel}}e^{\tau(\beta-\frac{1}{t_{rel}})}-(\beta-\frac{1}{t_{rel}})e^{-\frac{2\tau}{t_{rel}}}}{(\beta+\frac{1}{t_{rel}})^2(\beta-\frac{1}{t_{rel}})^2}+\frac{\frac{1}{t_{rel}}e^{2\beta\tau}-\beta-\frac{1}{t_{rel}}+\beta e^{-\frac{2\tau}{t_{rel}}}}{2\beta(\beta+\frac{1}{t_{rel}})^2}-\frac{\frac{1}{t_{rel}}e^{2\tau(\beta-\frac{1}{t_{rel}})}-\beta+(\beta-\frac{1}{t_{rel}})e^{-\frac{2\tau}{t_{rel}}}}{2\beta(\beta-\frac{1}{t_{rel}})^2}\right), \quad \text{(A7)}$$

where we used the fact that $\langle x(t+t')x(t+t'')|x(t),x(t+\tau)\rangle$ is symmetric in $t'$ and $t''$. We recall that for functions with the symmetry $f(t',t'') = f(t'',t')$ it holds: $\int_0^\tau\int_0^\tau dt'dt''f(t',t'') = 2\int_0^\tau dt'\int_0^{t'}dt''f(t',t'')$.

The limits for small and large time intervals $\tau$ verify respectively $\lim_{\tau\to 0}\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)} = 0$ and $\lim_{\tau\to\infty}\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)} = \alpha^2\sigma_x^2\frac{t_{rel}}{\beta(\beta t_{rel}+1)^2} = \sigma^2_{y(t)|x(t)}$.

The factorization of the probability density $p(\zeta_\tau^{xy})$ into conditional densities (Equation (A2)) leads to a decomposition of the mapping irreversibility. Here we show that in the BLRM all these terms are zero except for the two terms corresponding to the full-conditional density of the evolution of the response in the original trajectory and in the time-reversed conjugate.

For a stationary stochastic process such as the BLRM it holds $p(x(t) = \gamma, y(t) = \delta) = p(x(t+\tau) = \gamma, y(t+\tau) = \delta)$, then these two terms cancel:

$$\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty d\gamma d\delta d\mu d\xi\, p(x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu, y(t+\tau) = \xi)\cdot\ln(p(x(t) = \gamma, y(t) = \delta)) =$$

$$= \int_{-\infty}^\infty\int_{-\infty}^\infty d\gamma d\delta\, p(x(t) = \gamma, y(t) = \delta)\cdot\ln(p(x(t) = \gamma, y(t) = \delta)) =$$

$$= \int_{-\infty}^\infty\int_{-\infty}^\infty d\gamma d\delta\, p(x(t+\tau) = \gamma, y(t+\tau) = \delta)\cdot\ln(p(x(t) = \gamma, y(t) = \delta)) =$$

$$= \int_{-\infty}^\infty\int_{-\infty}^\infty d\mu d\xi\, p(x(t+\tau) = \mu, y(t+\tau) = \xi)\cdot\ln(p(x(t) = \mu, y(t) = \xi)) =$$

$$= \int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty d\gamma d\delta d\mu d\xi\, p(x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu, y(t+\tau) = \xi)\cdot\ln(p(x(t) = \mu, y(t) = \xi)). \quad \text{(A8)}$$

The contribution from the signal in the mapping irreversibility is also zero since the Ornstein-Uhlenbeck process is reversible, $p(x(t) = \gamma, x(t+\tau) = \mu) = p(x(t) = \mu, x(t+\tau) = \gamma)$:

$$\int_{-\infty}^\infty\int_{-\infty}^\infty d\gamma d\mu\, p(x(t) = \gamma, x(t+\tau) = \mu)\ln\left(\frac{p(x(t+\tau) = \mu|x(t) = \gamma)}{p(x(t+\tau) = \gamma|x(t) = \mu)}\right) = 0. \quad \text{(A9)}$$

The mapping irreversibility is therefore:

$$\Phi_\tau^{xy} = \langle\varphi_\tau^{xy}\rangle_{p(\zeta_\tau^{xy})} = \int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty d\gamma d\delta d\mu d\xi\, p(x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu, y(t+\tau) = \xi)*$$

$$*\ln\left(\frac{p(y(t+\tau) = \xi|x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu)}{p(y(t+\tau) = \delta|x(t) = \mu, y(t) = \xi, x(t+\tau) = \gamma)}\right) =$$

$$= -\frac{1}{2} + \frac{1}{2\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)}}\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty d\gamma d\delta d\mu d\xi\, p(x(t) = \gamma, y(t) = \delta, x(t+\tau) = \mu, y(t+\tau) = \xi)*$$

$$*(\delta - \langle y(t+\tau)|x(t) = \mu, y(t) = \xi, x(t+\tau) = \gamma\rangle)^2, \quad \text{(A10)}$$

where in the last passage we exploited the fact that all the probability densities are Gaussian distributed. Solving the integrals, we get the mapping irreversibility for the BLRM as a function of the time interval $\tau$:

$$\Phi_\tau^{xy} = \frac{1}{2}(e^{-2\beta\tau} - 1) + \frac{2\alpha^2\sigma_x^2 t_{rel}^2}{\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)}}\frac{(e^{-\beta\tau} - e^{-\frac{\tau}{t_{rel}}})^2(e^{-2\beta\tau} + 1 - 2e^{-\tau(\beta+\frac{1}{t_{rel}})})}{(\beta^2 t_{rel}^2 - 1)^2(1 - e^{-\frac{2\tau}{t_{rel}}})}. \quad \text{(A11)}$$

$\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)}$ is proportional to $\alpha^2\sigma_x^2$, therefore the mapping irreversibility $\Phi_\tau^{xy}$ is a function of just the two parameters $t_{rel}$ and $\beta$ (and of the observational time $\tau$).
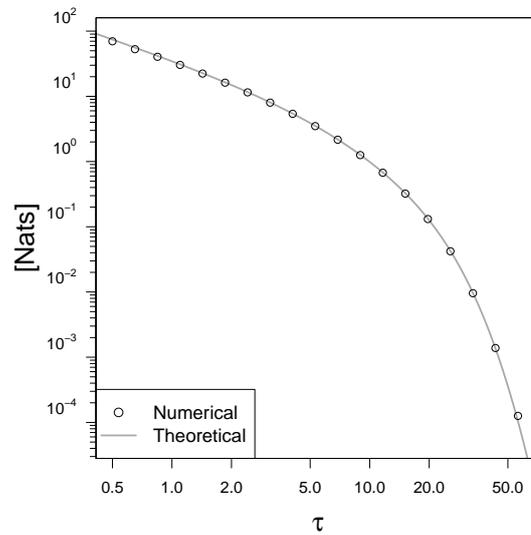
**Figure A1.** Numerical verification of the analytical solution for the entropy production $\Phi_\tau^{xy}$ with observational time $\tau$ in the BLRM. The parameters are $\beta = 0.2$ and $t_{rel} = 10$. The slight down-deviation for small $\tau$ is due to the finite box length in the discretized space, while the up-deviation for $\tau \to \infty$ is due to the finite number of samples.

## Appendix B. Backward Transfer Entropy in the BLRM

In the BLRM, where all densities are Gaussian distributed, the backward transfer entropy is equivalent to the time-reversed Granger causality [55]:

$$T_{y \to x}(-\tau) = I(x(t), y(t+\tau)|x(t+\tau)) = \ln\left(\frac{\sigma_{y|x}}{\sigma_{y(t+\tau)|x(t),x(t+\tau)}}\right) \tag{A12}$$

We must calculate the conditional variance $\sigma_{y(t+\tau)|x(t),x(t+\tau)}$. Let us recall the relation for the value of the response as a function of the whole history of the signal trajectory:

$$y(t+\tau) = \alpha e^{-\beta(t+\tau)}\left(\int_{-\infty}^{t} dt' x(t') e^{\beta t'} + \int_{t}^{t+\tau} dt' x(t') e^{\beta t'}\right) \tag{A13}$$

Then we write the conditional squared response as:

$$\langle y^2(t+\tau)|x(t), x(t+\tau)\rangle = 2\alpha^2 e^{-2\beta(t+\tau)}(e^{2\beta t}\int_0^\tau \int_0^{t''} dt''dt' e^{\beta(t'+t'')} \langle x(t+t')x(t+t'')|x(t), x(t+\tau)\rangle_{t'' \geq t'} +$$

$$+ \int_{-\infty}^{t}\int_{-\infty}^{t''} dt''dt' \langle x^2(t'')|x(t)\rangle e^{-\frac{t''-t'}{t_{rel}}+\beta(t'+t'')} + \int_{-\infty}^{t}\int_t^{t+\tau} dt'dt'' \langle x(t')|x(t)\rangle \langle x(t'')|x(t), x(t+\tau)\rangle e^{\beta(t'+t'')}). \tag{A14}$$

Since $\sigma^2_{y(t+\tau)|x(t),x(t+\tau)}$ is expected to be independent of $x(t)$ and $x(t+\tau)$, then the remaining terms after calculation sum up to:

$$\sigma^2_{y(t+\tau)|x(t),x(t+\tau)} = \sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)} + \sigma^2_{y|x} e^{-2\beta\tau}, \tag{A15}$$

where $\sigma^2_{y|x} = \frac{\sigma_x^2 \alpha^2}{\beta t_{rel}(\beta + \frac{1}{t_{rel}})^2}$ was already calculated in [30]. Then the backward transfer entropy is:

$$T_{y \to x}(-\tau) = I\left(x(t), y(t)|x(t+\tau)\right) = -\frac{1}{2} \ln \left( \frac{\sigma^2_{y(t+\tau)|x(t),y(t),x(t+\tau)}}{\sigma^2_{y|x}} + e^{-2\beta\tau} \right). \tag{A16}$$

### Appendix C. The Causal Influence Rate Converges to the Horowitz-Esposito Information Flow in the BLRM

We introduced the Horowitz-Esposito information flow [18,36] in Equation (3). In our stationary processes framework, the two components of the information flow are related by $I_{x \to y} = -I_{y \to x}$, so that the information flow is unidirectional and necessarily asymmetric when present. The $y$ variable in the BLRM is measuring the $x$ variable, therefore the information is flowing in the $x \to y$ direction, and we wish to calculate the parameter dependence of the positive $I^{x \to y}$:

$$I^{x \to y} = \int \int dx dy \, J_y(x,y) \frac{\frac{\partial}{\partial y} p(x|y)}{p(x|y)}. \tag{A17}$$

We see that the information flow is a function of probability currents. We plot the current $\vec{J}$ for the BLRM in Figure A2.
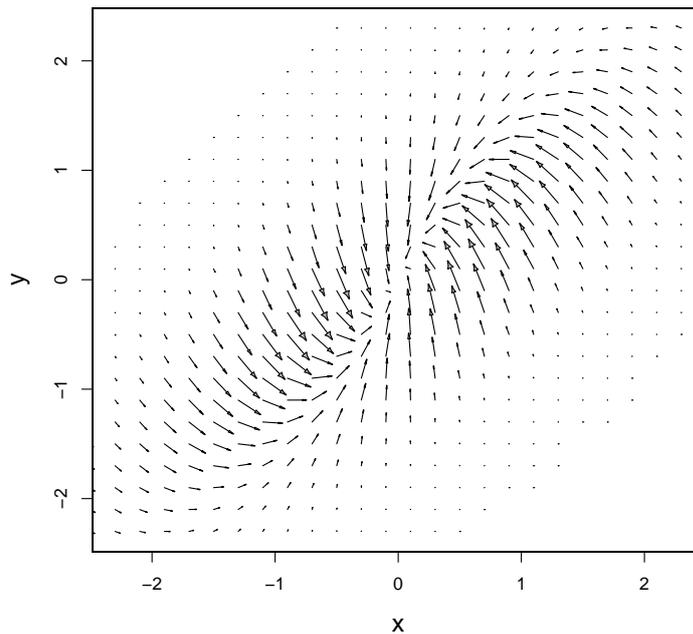


**Figure A2.** Probability currents $\vec{J}$ in the BLRM, estimated with $\tau = 0.1$. The parameters are $\beta = 0.2$ and $t_{rel} = 10$. The space coordinates are in units of the standard deviation.

The probability current $J_y(x,y)$ in the $y$ direction for the BLRM is given by $J_y = (\alpha x - \beta y)p(x,y)$. Then we calculate:

$$I^{x \to y} = \int \int dx dy \, p(y)(\alpha x - \beta y)\frac{\partial p(x|y)}{\partial y} = \beta - \int \int dx dy \, p(x|y)(\alpha x - \beta y)\frac{\partial p(y)}{\partial y} =$$
$$= \beta + \int \int dx dy (\alpha x - \beta y)p(x,y)\frac{y}{\sigma_y^2} = \frac{\alpha}{\sigma_y^2} \langle xy \rangle, \tag{A18}$$

where in the second passage we used partial integration and $p(y)(\alpha x - \beta y)$ is exponentially vanishing for $y \to \pm\infty$ because $p(y)$ is a Gaussian, $p(y) = N(0, \sigma_y^2)$. In the last passage we identified $\int \int dxdy \, p(x,y)xy = \langle xy \rangle$. Since the BLRM is a stationary process the time derivatives of expectations vanish, then $0 = \frac{d\langle x(t)y(t) \rangle}{dt} = \left\langle \frac{dx}{dt} y \right\rangle + \left\langle x \frac{dy}{dt} \right\rangle = -\langle xy \rangle (\beta + \frac{1}{t_{rel}}) + \alpha\sigma_x^2$, and we find $\langle xy \rangle = \frac{\alpha\sigma_x^2}{\beta + \frac{1}{t_{rel}}}$.

Then using the BLRM expression [30] for the variance of the response $\sigma_y^2 = \frac{\alpha^2 t_{rel}}{\beta(\beta t_{rel}+1)}\sigma_x^2$ we obtain:

$$I^{x \to y} = \beta. \tag{A19}$$

The Horowitz-Esposito information flow in the BLRM is equal to the inverse of the deterministic response time to perturbations $\frac{1}{\beta}$. Interestingly, this is independent of the time scale of fluctuations $t_{rel}$. Let us consider a fixed $\beta$, then if $t_{rel}$ is small we have very fast fluctuations and the response is not able to follow the signal with accuracy and the mutual information $I(x,y) = \frac{1}{2}\ln(1 + \beta t_{rel})$ is small. Nevertheless, the information flow $I^{x \to y}$ does not decrease because the dynamics of the $y$ variable is driven by the $x$ position for every possible situation $(x, y)$ even if not strongly correlated.

Importantly, in the small observational time limit our definition of causal influence [30] converges in rate to the Horowitz-Esposito information flow:

$$\lim_{\tau \to 0} \frac{C_{x \to y}(\tau)}{\tau} = I^{x \to y}. \tag{A20}$$

## Appendix D. Numerical Convergence of the Mapping Irreversibility to the Entropy Production in the Feedback Cooling Model

The feedback cooling model [19,36] describes a Brownian particle with velocity $x$ and viscous damping $\gamma$, that is under the feedback control of the measurement device $y$. The variable $y$ is a low-pass filter of noisy measurements on $x$. The SDE system describing the process is written:

$$\begin{cases} dx = -(\gamma x + ky) \, dt + \sqrt{D_x} \, dW_x \\ dy = (x - y) \, dt + \sqrt{D_y} \, dW_y \end{cases} \tag{A21}$$

where $k > 0$ is the feedback intensity. The mapping irreversibility (8) converges in the limit of small observational time $\tau \to 0$ to the physical entropy production (5) if the conditional probability $p(x_{t+\tau}, y_{t+\tau}|x_t, y_t) = p(x_{t+\tau}|x_t, y_t) \cdot p(y_{t+\tau}|x_t, y_t, x_{t+\tau})$ converges almost surely to the bipartite form $p(x_{t+\tau}|x_t, y_t) \cdot p(y_{t+\tau}|x_t, y_t)$. Importantly, the convergence has to be faster than $\tau$ so that in the limit of continuous trajectories we can almost surely neglect the term $\lim_{\tau \to 0} \frac{1}{\tau} \ln \frac{p(y_{t+\tau}|x_t, y_t, x_{t+\tau})}{p(y_{t+\tau}|x_t, y_t)} = 0$.

The knowledge of $x_{t+\tau}$ acts only on the estimate of $f_y(x_{t+t'}, y_{t+t'})$ (with $0 \le t' \le \tau$) because the diffusion coefficients are constant. Since the system (A21) is linear, the Kullback–Leibler divergence can be expressed in terms of conditional expectations:

$$\left\langle \ln \frac{p(y_{t+\tau}|x_t, y_t, x_{t+\tau})}{p(y_{t+\tau}|x_t, y_t)} \right\rangle_{p(\zeta_\tau^{xy})} =$$

$$= -\ln \frac{\sigma_{y_{t+\tau}|x_t, y_t, x_{t+\tau}}}{\sigma_{y_{t+\tau}|x_t, y_t}} + \frac{\sigma_{y_{t+\tau}|x_t, y_t, x_{t+\tau}}^2 + \left\langle ((y_{t+\tau}|x_t, y_t, x_{t+\tau}) - \langle y_{t+\tau}|x_t, y_t\rangle)^2 \right\rangle_{p(x_t, y_t, x_{t+\tau})}}{2\sigma_{y_{t+\tau}|x_t, y_t}^2} - \frac{1}{2}. \tag{A22}$$

The conditional variance $\sigma_{y_{t+\tau}|x_t, y_t, x_{t+\tau}}^2$ is of order $\langle W_y^2(\tau) \rangle \sim \tau$ and differs from $\sigma_{y_{t+\tau}|x_t, y_t}^2$ only with a term of order $(\partial_x f_y(x,y))^2 \langle W_x^2(\tau) \rangle \tau^2 \sim \tau^3$ so that $\ln \frac{\sigma_{y_{t+\tau}|x_t, y_t, x_{t+\tau}}}{\sigma_{y_{t+\tau}|x_t, y_t}} \sim \tau^2$ for $\tau \to 0$.

While the conditional variances are constant, the conditional expectation $\langle y_{t+\tau}|x_t, y_t, x_{t+\tau} \rangle$ depend linearly on $x_{t+\tau}$ (and on $x_t$, $y_t$), therefore it is sufficient to look at the conditional correlation $C(x_{t+\tau}, y_{t+\tau}|x_t, y_t) = \frac{\langle x_{t+\tau} y_{t+\tau}|x_t, y_t \rangle - \langle x_{t+\tau}|x_t, y_t \rangle \langle y_{t+\tau}|x_t, y_t \rangle}{\sigma_{x_{t+\tau}|x_t, y_t} \sigma_{y_{t+\tau}|x_t, y_t}}$, given that

$$\left\langle \frac{(\langle y_{t+\tau}|x_t,y_t,x_{t+\tau}\rangle - \langle y_{t+\tau}|x_t,y_t\rangle)^2}{\sigma_{y_{t+\tau}|x_t,y_t}} \right\rangle_{p(x_{t+\tau}|x_t,y_t)} = C^2(x_{t+\tau},y_{t+\tau}|x_t,y_t).$$

By numerical simulation we checked that $\frac{1}{\tau}C^2(x_{t+\tau},y_{t+\tau}|x_t,y_t) \to 0$ in the limit $\tau \to 0$ for the feedback cooling model (A21). Importantly, we checked that with $D_y = 0$ there is no convergence, as it is also the case for the BLRM.

For the case of nonconstant diffusion coefficients (multiplicative noise) the argument on the conditional variances does not hold, and we are not sure of the convergence. The idea is that there should be an intermediate case between vanishing with constant $D > 0$ and diverging with $D = 0$, and this could be the case of multiplicative noise. Such characterization is beyond the scope of this paper.

## Appendix E. Numerical Estimation of the Entropy Production in the Bivariate Gaussian Approximation

We calculate numerically the mapping irreversibility $\Phi_\tau^{xy}$ as an average of the spatial density of irreversibility $\psi(x_t,y_t)$, $\Phi_\tau^{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_t dy_t \psi(x_t,y_t)$. In the computer algorithm the $(x,y)$ space is discretized in boxes $(i,j)$, and for each box we estimate the conditional correlation $C_{xy|i,j}$ of future values $(x_{t+\tau},y_{t+\tau})$, the conditional correlation $\widetilde{C}_{xy|i,j}$ of past values $(x_{t-\tau},y_{t-\tau})$, the expected values for both variables in future ($\langle x|i,j\rangle$, $\langle y|i,j\rangle$) and past states ($\langle \widetilde{x}|i,j\rangle$, $\langle \widetilde{y}|i,j\rangle$), and the standard deviations on those $\sigma_{x|i,j}$, $\widetilde{\sigma}_{x|i,j}$, $\sigma_{y|i,j}$, $\widetilde{\sigma}_{y|i,j}$. The spatial density evaluated in the box $(i,j)$ is then calculated as the bidimensional Kullback–Leibler divergence in the Gaussian approximation [56]:

$$\psi(i,j) = P(i,j)\,\frac{1}{2}\Big[\ln\Big(\frac{\widetilde{\sigma}_{x|i,j}^2 \widetilde{\sigma}_{y|i,j}^2(1-\widetilde{C}_{xy|i,j}^2)}{\sigma_{x|i,j}^2 \sigma_{y|i,j}^2(1-C_{xy|i,j}^2)}\Big) - 2 + \frac{\frac{\sigma_{x|i,j}^2}{\widetilde{\sigma}_{x|i,j}^2} + \frac{\sigma_{y|i,j}^2}{\widetilde{\sigma}_{y|i,j}^2} - 2\frac{\sigma_{x|i,j}\sigma_{y|i,j}}{\widetilde{\sigma}_{x|i,j}\widetilde{\sigma}_{y|i,j}}C_{xy|i,j}\widetilde{C}_{xy|i,j}}{1-\widetilde{C}_{xy|i,j}^2} +$$

$$+ \frac{\widetilde{\sigma}_{y|i,j}^2(\langle \widetilde{x}|i,j\rangle - \langle x|i,j\rangle)^2 + \widetilde{\sigma}_{x|i,j}^2(\langle \widetilde{y}|i,j\rangle - \langle y|i,j\rangle)^2 - 2\widetilde{C}_{xy|i,j}\widetilde{\sigma}_{x|i,j}\widetilde{\sigma}_{y|i,j}(\langle \widetilde{x}|i,j\rangle - \langle x|i,j\rangle)(\langle \widetilde{y}|i,j\rangle - \langle y|i,j\rangle)}{\widetilde{\sigma}_{x|i,j}^2 \widetilde{\sigma}_{y|i,j}^2(1-\widetilde{C}_{xy|i,j}^2)}\Big] \tag{A23}$$

The effect of the finite width of the discretization is attenuated by estimating all the quantities taking into account the starting point $(x_t,y_t)$ within the box $(i,j)$, subtracting the difference to the mean values for each box. For example, when we sample for the estimate of the conditional average $\langle x_{t+\tau}|i\rangle$ we would collect samples $x_{t+\tau} - (x_t - \langle x_t|i\rangle)$.

## References

1. Jarzynski, C. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 329–351. [CrossRef]
2. Parrondo, J.M.; van den Broeck, C.; Kawai, R. Entropy production and the arrow of time. *New J. Phys.* **2009**, *11*, 073008. [CrossRef]
3. Feng, E.H.; Crooks, G.E. Length of time's arrow. *Phys. Revi. Lett.* **2008**, *101*, 090602. [CrossRef] [PubMed]
4. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690. [CrossRef]
5. Crooks, G.E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **1999**, *60*, 2721. [CrossRef]
6. Evans, D.J.; Searles, D.J. The fluctuation theorem. *Adv. Phys.* **2002**, *51*, 1529–1585. [CrossRef]
7. Kawai, R.; Parrondo, J.; van den Broeck, C. Dissipation: The phase-space perspective. *Phys. Rev. Lett.* **2007**, *98*, 080602. [CrossRef] [PubMed]
8. Jarzynski, C. Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Phys.* **2000**, *98*, 77–102. [CrossRef]
9. Chernyak, V.Y.; Chertkov, M.; Jarzynski, C. Path-integral analysis of fluctuation theorems for general Langevin processes. *J. Stat. Mech. Theory Exp.* **2006**, *2006*, P08001. [CrossRef]
10. Ito, S.; Sagawa, T. Information thermodynamics on causal networks. *Phys. Rev. Lett.* **2013**, *111*, 180603. [CrossRef]
11. Sagawa, T.; Ueda, M. Nonequilibrium thermodynamics of feedback control. *Phys. Rev. E* **2012**, *85*, 021104. [CrossRef]

12. Sagawa, T.; Ueda, M. Generalized Jarzynski equality under nonequilibrium feedback control. *Phys. Rev. Lett.* **2010**, *104*, 090602. [CrossRef]

13. Szilard, L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Syst. Res. Behav. Sci.* **1964**, *9*, 301–310. [CrossRef]

14. Martínez, I.A.; Roldán, É.; Dinis, L.; Petrov, D.; Parrondo, J.M.; Rica, R.A. Brownian carnot engine. *Nat. Phys.* **2016**, *12*, 67–70. [CrossRef] [PubMed]

15. Ciliberto, S. Experiments in stochastic thermodynamics: Short history and perspectives. *Phys. Rev. X* **2017**, *7*, 021051. [CrossRef]

16. Toyabe, S.; Sagawa, T.; Ueda, M.; Muneyuki, E.; Sano, M. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* **2010**, *6*, 988–992. [CrossRef]

17. Koski, J.V.; Maisi, V.F.; Pekola, J.P.; Averin, D.V. Experimental realization of a Szilard engine with a single electron. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13786–13789. [CrossRef] [PubMed]

18. Horowitz, J.M.; Esposito, M. Thermodynamics with continuous information flow. *Phys. Rev. X* **2014**, *4*, 031015. [CrossRef]

19. Rosinberg, M.L.; Horowitz, J.M. Continuous information flow fluctuations. *EPL (Europhysics Letters)* **2016**, *116*, 10007. [CrossRef]

20. Horowitz, J.M. Multipartite information flow for multiple Maxwell demons. *J. Stat. Mech. Theory Exp.* **2015**, *2015*, P03006. [CrossRef]

21. Ito, S. Backward transfer entropy: Informational measure for detecting hidden Markov models and its interpretations in thermodynamics, gambling and causality. *Sci. Rep.* **2016**, *6*. [CrossRef] [PubMed]

22. Spinney, R.E.; Lizier, J.T.; Prokopenko, M. Transfer entropy in physical systems and the arrow of time. *Phys. Rev. E* **2016**, *94*, 022135. [CrossRef] [PubMed]

23. Parrondo, J.M.; Horowitz, J.M.; Sagawa, T. Thermodynamics of information. *Nat. Phys.* **2015**, *11*, 131–139. [CrossRef]

24. Ito, S. Unified framework for the second law of thermodynamics and information thermodynamics based on information geometry. *arXiv* **2018**, arXiv:1810.09545

25. Ito, S. Stochastic thermodynamic interpretation of information geometry. *Phys. Rev. Lett.* **2018**, *121*, 030605. [CrossRef] [PubMed]

26. Roldán, É.; Parrondo, J.M. Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems. *Phys. Rev. E* **2012**, *85*, 031129. [CrossRef] [PubMed]

27. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.

28. Sekimoto, K. Langevin equation and thermodynamics. *Prog. Theor. Phys. Supp.* **1998**, *130*, 17–27. [CrossRef]

29. Crooks, G.E.; Still, S.E. Marginal and conditional second laws of thermodynamics. *arXiv* **2016**, arXiv:1611.04628.

30. Auconi, A.; Giansanti, A.; Klipp, E. Causal influence in linear Langevin networks without feedback. *Phys. Rev. E* **2017**, *95*, 042315. [CrossRef]

31. Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* **2012**, *75*, 126001. [CrossRef]

32. Shreve, S.E. *Stochastic Calculus for Finance II: Continuous-Time Models*; Springer Science & Business Media: New York, NY, USA, 2004; Volume 11.

33. Taniguchi, T.; Cohen, E. Onsager-Machlup theory for nonequilibrium steady states and fluctuation theorems. *J. Stat. Phys.* **2007**, *126*, 1–41. [CrossRef]

34. Seifert, U. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys. Rev. Lett.* **2005**, *95*, 040602. [CrossRef]

35. Risken, H. Fokker-planck equation. In *The Fokker-Planck Equation*; Springer: Berlin, Germany, 1996; pp. 63–95.

36. Horowitz, J.M.; Sandberg, H. Second-law-like inequalities with information and their interpretations. *New J. Phys.* **2014**, *16*, 125007. [CrossRef]

37. Gomez-Marin, A.; Parrondo, J.M.; van den Broeck, C. Lower bounds on dissipation upon coarse graining. *Phys. Rev. E* **2008**, *78*, 011107. [CrossRef] [PubMed]

38. Ito, S.; Sagawa, T. Information flow and entropy production on Bayesian networks. In *Mathematical Foundations and Applications of Graph Entropy*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2016; pp. 63–100.

39. Uhlenbeck, G.E.; Ornstein, L.S. On the theory of the Brownian motion. *Phys. Rev.* **1930**, *36*, 823. [CrossRef]

40. Gillespie, D.T. Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Phys. Rev. E* **1996**, *54*, 2084. [CrossRef]

41. R Core Team. *R: A Language and Environment for Statistical Computing*; Foundation for Statistical Computing, Vienna, Austria, 2014. Available online: www.r-project.org (accessed on 14 Febuary 2019).

42. Barrett, A.B. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E* **2015**, *91*, 052802. [CrossRef] [PubMed]

43. Klipp, E.; Liebermeister, W.; Wierling, C.; Kowald, A.; Herwig, R. *Systems Biology: A Textbook*; John Wiley & Sons; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2016.

44. Kholodenko, B.N. Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 165. [CrossRef] [PubMed]

45. Bialek, W.; Setayeshgar, S. Physical limits to biochemical signaling. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10040–10045. [CrossRef]

46. Tkačik, G.; Callan, C.G.; Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12265–12270. [CrossRef]

47. Crisanti, A.; de Martino, A.; Fiorentino, J. Statistics of optimal information flow in ensembles of regulatory motifs. *Phys. Rev. E* **2018**, *97*, 022407. [CrossRef]

48. Waltermann, C.; Klipp, E. Information theory based approaches to cellular signaling. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2011**, *1810*, 924–932. [CrossRef]

49. Di Talia, S.; Wieschaus, E.F. Short-term integration of Cdc25 dynamics controls mitotic entry during Drosophila gastrulation. *Dev. Cell* **2012**, *22*, 763–774. [CrossRef] [PubMed]

50. Nemenman, I. Gain control in molecular information processing: Lessons from neuroscience. *Phys. Biol.* **2012**, *9*, 026003. [CrossRef] [PubMed]

51. Di Talia, S.; Wieschaus, E.F. Simple biochemical pathways far from steady state can provide switchlike and integrated responses. *Biophys. J.* **2014**, *107*, L1–L4. [CrossRef] [PubMed]

52. Tkačik, G.; Walczak, A.M.; Bialek, W. Optimizing information flow in small genetic networks. *Phys. Rev. E* **2009**, *80*, 031920. [CrossRef] [PubMed]

53. Jarzynski, C. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E* **2006**, *73*, 046105. [CrossRef] [PubMed]

54. Ito, S.; Sagawa, T. Maxwell's demon in biochemical signal transduction with feedback loop. *Nat. Commun.* **2015**, *6*, 7498. [CrossRef] [PubMed]

55. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [CrossRef]

56. Duchi, J. Derivations for Linear Algebra and Optimization. Available online: https://docplayer.net/30887339-Derivations-for-linear-algebra-and-optimization.html (accessed on 11 Febuary 2019).