

# Beyond thermodynamic uncertainty relations: nonlinear response, error-dissipation trade-offs, and speed limits

Gianmaria Falasco<sup>1,\*</sup> , Massimiliano Esposito<sup>1</sup> and Jean-Charles Delvenne<sup>2</sup>

<sup>1</sup> Complex Systems and Statistical Mechanics, Physics and Materials Science Department, University of Luxembourg, L-1511, Luxembourg

<sup>2</sup> Institute of Information and Communication Technologies, Electronics and Applied Mathematics Université catholique de Louvain, Louvain-La-Neuve, Belgium

E-mail: [gianmaria.falasco@uni.lu](mailto:gianmaria.falasco@uni.lu), [massimiliano.esposito@uni.lu](mailto:massimiliano.esposito@uni.lu) and [jean-charles.delvenne@uclouvain.be](mailto:jean-charles.delvenne@uclouvain.be)

Received 24 September 2021, revised 7 February 2022

Accepted for publication 8 February 2022

Published 24 February 2022



## Abstract

From a recent geometric generalization of thermodynamic uncertainty relations (TURs) we derive novel upper bounds on the nonlinear response of an observable of an arbitrary system undergoing a change of probabilistic state. Various relaxations of these bounds allow to recover well known bounds such as (strengthenings of) Cramer–Rao’s and Pinsker’s inequalities. In particular we obtain a master inequality, named symmetric response intensity relation, which recovers several TURs as particular cases. We employ this set of bounds for three physical applications. First, we derive a trade-off between thermodynamic cost (dissipated free energy) and reliability of systems switching instantly between two states, such as one-bit memories. We derive in particular a lower bound of  $2.8k_B T$  per Shannon bit to write a bit in such a memory, a bound distinct from Landauer’s one. Second, we obtain a new family of classic speed limits which provide lower bounds for non-autonomous Markov processes on the time needed to transition between two probabilistic states in terms of a thermodynamic quantity (e.g. non-equilibrium free energy) and a kinetic quantity (e.g. dynamical activity). Third, we provide an upper bound on the nonlinear response of a system based solely on the ‘complexity’ of the system (which we

\*Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

relate to a high entropy and non-uniformity of the probabilities). We find that ‘complex’ models (e.g. with many states) are necessarily fragile to some perturbations, while simple systems are robust, in that they display a low response to arbitrary perturbations.

Keywords: stochastic thermodynamics, thermodynamic uncertainty relations, response theory, speed limits, thermodynamic cost of computation, nonequilibrium statistical mechanics

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Thermodynamic uncertainty relations (TURs) are as old as their better known counterparts of quantum mechanics. For more than a half century the term has denoted the tradeoff between the uncertainty in pairs of conjugated thermodynamic variables in systems at equilibrium (see [1] and references therein). Despite a long debate about their exact practical meaning, their derivation has been repeated several times, resorting either to the theory of equilibrium fluctuations or to purely statistical methods [2].

However, in the last few years the term has started to designate a radically different set of inequalities valid for nonequilibrium systems (see [3–14] for an incomplete list). Roughly speaking, TURs state that the mean-to-variance ratio of current-like observables is bounded by a function of the entropy production [15]. These TURs have been obtained by analogous methods as those used for the ancient thermodynamics uncertainty relations, in their modern forms now well known to physicists—large deviations theory [4, 6] and information theory [16]—and applied to rather general dynamics—Markov processes equipped by the thermodynamic notion of local detailed balance [17]. Recently, we have put forward an alternative derivation which deduces TURs from the very mathematical properties of the observable space [14]—somewhat similar to viewing the Heisenberg principle as a direct result of the uncertainty between Fourier transformed pairs of variables.

The fact that nonequilibrium conditions bring time and dynamics to the forefront has led the community to consider other bounds on dissipative processes. On the one hand, classical speed limits (SLs) have been derived which prescribe the minimal time to implement a system’s transformation in terms of the involved dissipation [18]. It is by now understood that TURs and SLs are just two faces of the same fundamental fact underlying all nonequilibrium conditions [19], that is the breaking of time-reversal invariance and hence the appearance of a non-zero average entropy production. On the other hand, results have started to appear on the response of systems to environmental perturbations [20]. The general picture emerging from these various results is the identification of the fundamental physical tradeoffs between efficiency, performance and robustness of nonequilibrium processes, as used in natural as well as man-made applications.

In this paper we contribute to the quest for a unifying theory. First we define the asymmetric response intensity of an observable  $f$  subject to a change of probability distribution from  $q$  to  $p$ . We recover the linear regime bound of [16] but also a nonlinear version, which turns out to coincide with a well-known generalization of Cramer–Rao’s relation. Then we introduce the symmetric response intensity, a related quantity that nevertheless obeys different inequalities. In particular we derive the symmetric response intensity relation, that contains the various TURs for anti-symmetric observables [10, 11, 21]. Remarkably, it allows to derive a relation between reliability and thermodynamic cost for memories or gates that relax abruptly from a

zero to a one conversely (thus dissipating entropy on the way). This model is not universal (as one can escape this relation by a slow protocol for instance) but is relevant for many current technologies using switched systems, even beyond computing devices. As a further application, we also recover a result on distinguishing the arrow of time, completing [22].

The maximum response intensities between two probability distributions can serve as a measure of distance or divergence between these distributions. We leverage this to obtain novel SLs [18, 19] for finite state Markov chains, which offer lower bounds on the time needed to reach a probability distribution from another, in terms of the distance of these probability distribution, the distance to stationarity and the maximum escape (or entrance) rates.

Finally we observe that the largest response one can obtain away from a probability distribution is still relatively small if this probability is ‘simple’, i.e. of low entropy and close to uniform distribution. Thus ‘simple’ systems are ‘robust’. In particular, one should be cautious in using simplistic toy models to model complex situations (a practice often found in biophysics), as these toy models may be much more robust to perturbations than the original systems.

To guide the reader though the text we collected the main results in the graph figure 1 and the most important quantities in table 1.

## 2. How to measure nonlinear response

Consider two possible probability distributions  $p$  and  $q$  on an arbitrary space  $\Omega$ . A given physical observable, i.e. a real random variable  $f : \Omega \rightarrow \mathbb{R}$  may exhibit two different expected values  $\langle f \rangle_p$  and  $\langle f \rangle_q$ . In order to check how significant the difference between the two is, we may compare it to the average fluctuations of the observable under  $p$  or  $q$ . Taking  $q$  as a ‘reference’, or ‘unperturbed’ probability distribution, and  $p$  as the ‘perturbed’ distribution, we see the following adimensional ratio as the *response intensity* of the observable  $f$  to the perturbation  $p$  of  $q$ :

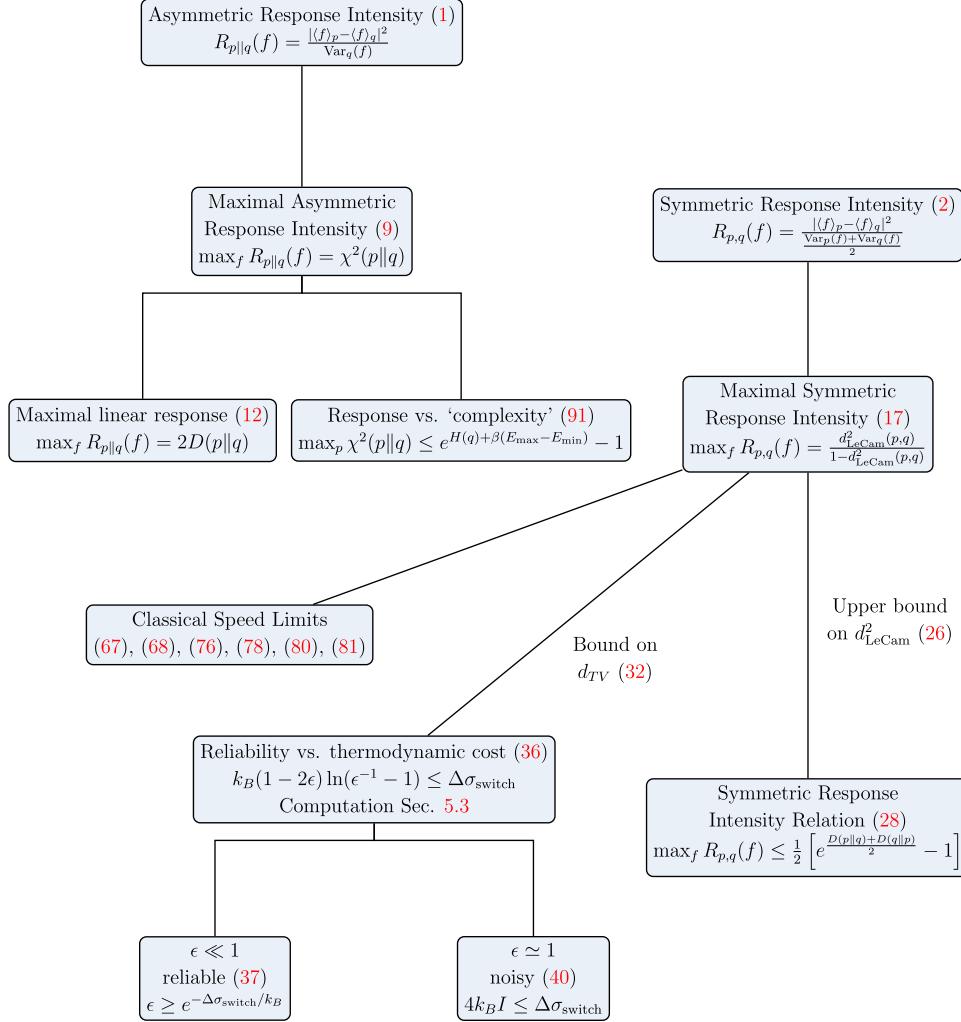
$$R_{p||q}(f) = \frac{|\langle f \rangle_p - \langle f \rangle_q|^2}{\text{Var}_q(f)}. \quad (1)$$

We call it the *asymmetric* response intensity as  $q$  and  $p$  have distinct roles. Sometimes  $p$  and  $q$  are two probability distributions of interest to describe a situation, without hierarchy of role. For example, think at two metastable states realizable within the same physical system. In this situation we may simply consider the average fluctuations over  $p$  and  $q$ , and define the *symmetric response intensity*:

$$R_{p,q}(f) = \frac{|\langle f \rangle_p - \langle f \rangle_q|^2}{\frac{\text{Var}_p(f) + \text{Var}_q(f)}{2}}. \quad (2)$$

One may intuitively see these quantities as how easy it is to tell  $p$  and  $q$  apart by just observing the value of  $f$ . If the response intensity  $R_{p||q}(f)$  is less than one, then it means that the typical values taken by  $f$  under  $p$  or under  $q$  differ by less than the typical fluctuation of  $f$ , hence it is hard to conclude whether the system is in the distribution  $p$  or  $q$ . Conversely, a large intensity means that the difference in response can hardly be confused with a random fluctuation.

The response intensity can be likened to a signal-to-noise ratio, if we see  $\langle f \rangle_p$  or  $\langle f \rangle_q$  as the signal being transmitted to the environment by a system switching between distributions  $p$  and  $q$ . This analogy can be made formal in the following scenario: if we choose to apply  $p$  or  $q$  randomly (with equal probabilities), then the signal, taking randomly the value  $\langle f \rangle_p$  or  $\langle f \rangle_q$ ,



**Figure 1.** Graphic outline of the main results of the paper and their logical connection.

**Table 1.** Summary of the main quantities used in the paper.

Kullback–Leibler divergence	$D(p  q) = \sum_{\omega \in \Omega} p(\omega) \log \frac{p(\omega)}{q(\omega)}$
$\chi^2$ divergence	$\chi^2(p  q) = \sum_{\omega \in \Omega} \frac{p^2(\omega)}{q(\omega)} - 1$
Le Cam’s distance	$d_{\text{LeCam}}(p, q) = \sqrt{\frac{1}{2} \sum_{\omega \in \Omega} \frac{(p(\omega) - q(\omega))^2}{p(\omega) + q(\omega)}}$
Shannon entropy	$H(q) = -\sum_{\omega \in \Omega} p(\omega) \log p(\omega)$
Total variation distance	$d_{\text{TV}}(p, q) = \frac{1}{2} \ p - q\ _1 = \frac{1}{2} \sum_{\omega \in \Omega}  p(\omega) - q(\omega) $

has a variance  $|\langle f \rangle_p - \langle f \rangle_q|^2/4$ , thus a signal-to-noise ratio (which is defined in information theory as the variance of the signal over the variance of the noise)  $R_{p,q}(f)/4$ .

### 3. Link between symmetric and asymmetric response intensities

We make some elementary observations on the response intensities introduced above. The symmetric response intensity is the harmonic average of asymmetric response intensities:

$$\frac{2}{R_{p,q}(f)} = \frac{1}{R_{p\parallel q}(f)} + \frac{1}{R_{q\parallel p}(f)}. \quad (3)$$

A key observation is that adding a constant to the observable  $f$  does not change  $R_{p\parallel q}(f)$  or  $R_{p,q}(f)$ , since both the response  $\langle f \rangle_p - \langle f \rangle_q$  and the variances are unaffected. Thus we can assume without loss of generality that  $\langle f \rangle_p = -\langle f \rangle_q$ . We call such an observable *centered*. Let us introduce an intermediate quantity where the variance is replaced with the second moment:

$$R_{p,q}^0(f) = \frac{|\langle f \rangle_p - \langle f \rangle_q|^2}{\frac{\langle |f|^2 \rangle_p + \langle |f|^2 \rangle_q}{2}}. \quad (4)$$

For a centered observable  $f$ , we find by direct calculation that

$$R_{p,q}(f)/4 = \frac{R_{p,q}^0(f)/4}{1 - R_{p,q}^0(f)/4} = \frac{1}{\frac{1}{R_{p,q}^0(f)/4} - 1} \quad (5)$$

and that

$$R_{p,q}^0(f)/4 = R_{p\parallel \frac{p+q}{2}}(f) = R_{q\parallel \frac{p+q}{2}}(f). \quad (6)$$

We would like now to find a tight upper bound on the response intensities  $R_{p,q}(f)$  and  $R_{p\parallel q}(f)$ , which would hold for all  $f$ , only depending on  $p$  and  $q$ .

### 4. Tight bounds on nonlinear response intensity and uncertainty principles

We exploit here the similar form of the response intensity and the square-mean-to-variance ratio appearing in TURs to derive bounds on the former in terms of physically interpretable quantities (such as the Kullback–Leibler divergences) between  $p$  and  $q$ . Specifically we use a recent geometric formalism introduced under the name of Hilbert uncertainty relation [14] which allows to prove general inequalities including, but not limited to, TURs, as we now recall.

#### 4.1. The Hilbert uncertainty relation: general statement

We consider a Hilbert space  $\mathcal{H}$  of observables, endowed with an arbitrary scalar product  $\langle \cdot | \cdot \rangle$ . One should not confuse  $\langle \cdot | \cdot \rangle$ , a standard notation for a scalar product, with  $\langle \cdot \rangle$ , a standard notation introduced above for the mean value (with respect to some probability distribution). Given a real continuous linear form  $f \mapsto L(f) \in \mathbb{R}$ , we want to find the maximum value of the ratio

$$\frac{|L(f)|^2}{\langle f | f \rangle} \quad (7)$$

among all  $f$  belonging to a given closed linear subspace  $\mathcal{F} \subseteq \mathcal{H}$  of observables of interest. We find that it reaches the maximum value for  $f = m \in \mathcal{F}$  such that  $L(g) = \langle m | g \rangle$  for all  $g \in \mathcal{F}$ ,

with  $m$  existing by virtue of Riesz's representation theorem. Then the maximum ratio takes the value

$$\max_f \frac{|L(f)|^2}{\langle f|f \rangle} = \langle m|m \rangle = |L(m)|. \quad (8)$$

Hereafter the Hilbert  $\mathcal{H}$  space can be finite or infinite-dimensional. As applications II and (part of) III below regard discrete systems, we spell out the general theory by using discrete summation notations, even though all results can be extended in straightforward way to continuous probability distributions, where the sums must in general be replaced with integrals.

#### 4.2. The Hilbert uncertainty relation: application to asymmetric response intensity

We can immediately apply the Hilbert uncertainty relation to derive the observables with maximal response intensity.

Let us start with the asymmetric case: given probability distributions  $p$  and  $q$ , we want to find  $f$  that maximises  $R_{p\|q}(f)$ . To apply the Hilbert uncertainty relation, we consider the linear form  $L(f) = \langle f \rangle_p - \langle f \rangle_q$  and the scalar product  $\langle f|g \rangle = \langle fg \rangle_q$  (the expected product  $fg$  according to distribution  $q$ ). Since, as observed above, we can always shift an observable by a constant and preserve the asymmetric response intensity  $R_{p\|q}(f)$ , we limit ourselves to the observables with zero  $q$ -mean:  $\mathcal{F} = \{f : \Omega \rightarrow \mathbb{R} | \langle f \rangle_q = 0\}$ . In this space, variance and second moment coincide, thus the asymmetric response intensity (1) is indeed the ratio (7). We easily find that the optimal observable is  $m = \frac{p}{q} - 1$ . Indeed we can check that  $\langle m \rangle_q = \sum_{\omega \in \Omega} q(\omega) \left( \frac{p(\omega)}{q(\omega)} - 1 \right) = 0$  and  $\langle m|g \rangle = \sum_{\omega \in \Omega} q(\omega) \left( \frac{p}{q} - 1 \right) g = L(g)$ .

Thus we conclude

$$\max_f R_{p\|q}(f) = \sum_{\omega \in \Omega} \frac{p^2}{q} - 1 = \chi^2(p\|q), \quad (9)$$

where  $\chi^2(p\|q)$  is the so-called  $\chi^2$  divergence of  $p$  with respect to  $q$ . Equation (9) formally coincides with the *Hammersley–Chapman–Robbins bound* [23] but has the merit of giving physical meaning to a result of estimation theory, by identifying in the left-hand side of (9) a sound proxy for the relative response intensity of a physical system. The latter bound is a known generalisation of Cramer–Rao's bound [23], which is obtained from *Hammersley–Chapman–Robbins bound* by considering  $p$  close to  $q$ , such that the  $\chi^2$  divergence reduces to Fisher's information metric.

The  $\chi^2$  divergence has the following relation with the Kullback–Leibler divergence, resulting from convexity of exponential:

$$\chi^2(p\|q) \geq e^{D(p\|q)} - 1 \geq D(p\|q). \quad (10)$$

This inequality can be interpreted physically in the following way. If we consider  $q$  as a stationary distribution of some Markov process and  $p$  as a non-stationary distribution relaxing to  $q$ , then  $k_B D(p\|q)$  is the entropy produced in the course of the relaxation, not counting the 'housekeeping' entropy possibly produced to maintain the stationary state  $q$ , if it is not an equilibrium distribution [24] (see section 5 for a more thorough discussion). This entropy production corresponds exactly to the *non-adiabatic entropy production* [25] for the case of autonomously relaxing systems. In the case where the stationary state  $q$  is an equilibrium at temperature  $T$ , then  $D(p\|q)k_B T$  is the non-equilibrium free energy being dissipated in the course of relaxation [26–28].

Thus if  $p$  is far from the stationary distribution  $q$  (in the sense of a large  $D(p\|q)$ ), then it results in the existence of a large nonlinear response intensity  $R_{p\|q}(f)$  for *some* observable  $f$ . In the limit where  $p$  and  $q$  are close (the so-called *linear response regime*), a Taylor expansion yields

$$\chi^2(p\|q) \approx \sum_{\omega \in \Omega} \frac{(p - q)^2}{q} \approx 2D(p\|q). \quad (11)$$

In summary, we have in the limit of linear regime:

$$\max_f R_{p\|q}(f) = 2D(p\|q) \quad (12)$$

which provides another proof of Dechant–Sasa’s linear-regime relation [16]. We notice that (12) is also the starting point to derive bounds on the mean-to-variance ratio of time-integrated observables with Markovian dynamics. This method, devised in [16], relies on an artificial perturbation which speeds up the dynamics. It can be implemented by adding auxiliary forces, or by appropriately rescaling the transition rates in jump dynamics. The two approaches yield, respectively, the mean entropy production [13, 29] and the mean dynamical activity [30] as the upper bound to the mean-to-variance ratio.

#### 4.3. The Hilbert uncertainty relation: application to symmetric response intensity

As we seek now to bound the symmetric response intensity, we choose the scalar product  $\langle f|g \rangle = \frac{\langle fg \rangle_p + \langle fg \rangle_q}{2} = \langle fg \rangle_{\frac{p+q}{2}}$ . The space of observables  $\mathcal{F}$  of interest is now the set of all observables. The linear form is still  $L(f) = \langle f \rangle_p - \langle f \rangle_q$ . In this case we see that the ratio (7) is exactly  $R_{p,q}^0(f)$ , defined in (4). The maximum ratio is reached for the observable  $m = 2\frac{p-q}{p+q}$  in application of the Hilbert uncertainty relation, and  $\langle m|g \rangle = L(g)$  for any observable  $g$ . Thus we arrive at

$$\max_f R_{p,q}^0(f) = 2 \sum_{\omega \in \Omega} \frac{(p - q)^2}{p + q} = 4d_{\text{LeCam}}^2(p, q). \quad (13)$$

The latter expression involves Le Cam’s distance [31, 32] between two probability distributions  $p$  and  $q$ , defined as

$$d_{\text{LeCam}}(p, q) = \sqrt{\frac{1}{2} \sum_{\omega \in \Omega} \frac{(p - q)^2}{p + q}} \leq 1. \quad (14)$$

This quantity is indeed a proper distance for the space of probability distributions over  $\Omega$ , verifying the axioms of a metric (including, non-trivially, the triangle inequality) [31].

Note that  $m$  happens to be centered, namely  $\langle m \rangle_p = -\langle m \rangle_q$ , thus the maximum in (14) is also the maximum of all centered observables  $f$ :

$$\max_{f \text{centered}} R_{p,q}^0(f) = 2 \sum_{\omega \in \Omega} \frac{(p - q)^2}{p + q} = 4d_{\text{LeCam}}^2(p, q). \quad (15)$$

In view of (6), the same observable  $m = 2\frac{p-q}{p+q}$  is also maximal for  $R_{p\|\frac{p+q}{2}}(f)$  and for  $R_{p\|\frac{p+q}{2}}(f)$ —over *all* observables  $f$ , centered or not, since those quantities are invariant under

shift by a constant. Thus from the Hammersley–Chapman–Robbins bound (9), we find that

$$d_{\text{LeCam}}^2(p, q) = \chi^2 \left( p \left\| \frac{p+q}{2} \right\| \right) = \chi^2 \left( q \left\| \frac{p+q}{2} \right\| \right), \quad (16)$$

as can also be verified by direct calculation. Finally, using (5) and (6), the symmetric response intensity is found to be bounded (for all observables  $f$ , centered or not) as follows:

$$\max_f R_{p,q}(f) = \frac{\frac{1}{2} \sum_{\omega \in \Omega} \frac{(p-q)^2}{p+q}}{1 - \frac{1}{2} \sum_{\omega \in \Omega} \frac{(p-q)^2}{p+q}} \quad (17)$$

$$= \frac{\chi^2(p \left\| \frac{p+q}{2} \right\|)}{1 - \chi^2(p \left\| \frac{p+q}{2} \right\|)} \quad (18)$$

$$= \frac{d_{\text{LeCam}}^2(p, q)}{1 - d_{\text{LeCam}}^2(p, q)}. \quad (19)$$

Bounding the symmetric response intensity thus reduces to bounding Le Cam's distance, e.g. with respect to other quantities such Kullback–Leibler divergences, as we now proceed to do.

#### 4.4. Upper bound on Le Cam's distance and the symmetric response intensity relation

Upper and lower bounds can be found on Le Cam's distance in terms of other distances between probability distributions (see appendix A for lower bounds). A well-known distance between probability distributions is the *total variation distance*

$$d_{\text{TV}}(p, q) = \frac{1}{2} \|p - q\|_1 = \frac{1}{2} \sum_{\omega \in \Omega} |p - q| = \sum_{\omega: p > q} (p - q) \leq 1. \quad (20)$$

We now derive a novel upper bound on Le Cam's distance based on the total variation distance and the Kullback–Leibler divergence. Given the distributions  $p, q$  on  $\Omega$ , we define  $\Omega_p = \{\omega \in \Omega \mid p(\omega) > q(\omega)\}$  and  $\Omega_q = \{\omega \in \Omega \mid q(\omega) \geq p(\omega)\}$ . In other words, we split  $\Omega$  into the part  $\Omega_p$  (resp.  $\Omega_q$ ) where  $p$  dominates  $q$  (resp. where  $q$  dominates  $p$ ). On  $\Omega_p$  we define the distribution  $p' = \frac{p-q}{d_{\text{TV}}(p, q)}$ , while on  $\Omega_q$  we define the distribution  $q' = \frac{q-p}{d_{\text{TV}}(p, q)}$ . We now observe that  $\frac{p-q}{p+q} = \tanh \frac{1}{2} \ln \frac{p}{q}$  and exploit the fact that  $\tanh$  is concave for positive arguments. In this way we find

$$\sum_{\omega \in \Omega} \frac{(p-q)^2}{p+q} = \sum_{\omega \in \Omega} (p-q) \tanh \left( \frac{1}{2} \ln \frac{p}{q} \right) \quad (21)$$

$$= d_{\text{TV}}(p, q) \left[ \sum_{\omega \in \Omega_p} p' \tanh \left( \frac{1}{2} \ln \frac{p}{q} \right) + \sum_{\omega \in \Omega_q} q' \tanh \left( \frac{1}{2} \ln \frac{q}{p} \right) \right] \quad (22)$$

$$\leq d_{\text{TV}}(p, q) \left[ \tanh \left( \frac{1}{2} \sum_{\omega \in \Omega_p} p' \ln \frac{p}{q} \right) + \tanh \left( \frac{1}{2} \sum_{\omega \in \Omega_q} q' \ln \frac{q}{p} \right) \right] \quad (23)$$

$$\leq 2d_{\text{TV}}(p, q) \tanh \left( \frac{1}{4d_{\text{TV}}(p, q)} \sum_{\omega \in \Omega} (p-q) \ln \frac{p}{q} \right) \quad (24)$$

$$= 2d_{\text{TV}}(p, q) \tanh \left( \frac{1}{2d_{\text{TV}}(p, q)} \frac{D(p\|q) + D(q\|p)}{2} \right). \quad (25)$$

To summarize, our first main result in this section is

$$d_{\text{LeCam}}^2(p, q) \leq d_{\text{TV}}(p, q) \tanh \frac{1}{2d_{\text{TV}}(p, q)} \frac{D(p\|q) + D(q\|p)}{2}. \quad (26)$$

Given that the rhs of (26) is an increasing function of  $d_{\text{TV}}(p, q)$ , we can relax it by setting  $d_{\text{TV}}(p, q)$  to its maximum value (one), and obtain a novel bound in terms of symmetric Kullback–Leibler divergence alone,

$$d_{\text{LeCam}}^2(p, q) \leq \tanh \frac{1}{2} \frac{D(p\|q) + D(q\|p)}{2} \quad (27)$$

and thus, via (17), the following bound on the response intensity,

$$\max_f R_{p,q}(f) \leq \frac{e^{\frac{D(p\|q)+D(q\|p)}{2}} - 1}{2}. \quad (28)$$

This is the second main result of this section, which we call the *symmetric response intensity relation*. Let us mention three immediate applications of these relations recovering previously known results.

First, taking  $\Omega$  as a set of trajectories for a stochastic system,  $p$  the probability distribution over trajectories and  $q$  the probability distribution over time-reversed trajectories, we obtain that  $D(p\|q) = D(q\|p)$ . We thus see that the symmetric response intensity relation (28) contains the various TURs derived in [10, 11, 21],

$$\frac{\langle f \rangle_p^2}{\text{Var}_p(f)} \leq \frac{e^{D(p\|q)} - 1}{2}. \quad (29)$$

The added generality is that the observables here need not be current-like observables, i.e. antisymmetric under time reversal. As pointed out in [14], these results hold more generally for any involution (order-two symmetry) over any probability space, and it turns out that this generality can be leveraged to provide another proof of (28), see appendix B. Moreover, [14] shows that the original TUR [3, 4, 6] can be derived as well from (29) under the additional assumptions of Markovianity and stationarity.

Second, by combining the lower bound and upper bounds on Le Cam's distance, we retrieve a useful bound on the Jensen–Shannon distance in terms of Kullback–Leibler divergences, see appendix C.

Third, an interesting observable to consider is  $f = \text{sign}(p - q)$ . Then we have  $\langle f \rangle_p - \langle f \rangle_q = 2d_{\text{TV}}(p, q)$  and  $\langle |f|^2 \rangle_p = \langle |f|^2 \rangle_q = 1$ , implying  $R_{p,q}^0(f) = 4d_{\text{TV}}^2(p, q)$ . Using this in (13) (which yields  $d_{\text{TV}} \leq d_{\text{LeCam}}$ ) and (26) allows us to find the following relation between total variation distance and Kullback–Leibler divergences,

$$d_{\text{TV}}(p, q) \leq \tanh \frac{1}{2d_{\text{TV}}(p, q)} \frac{D(p\|q) + D(q\|p)}{2}, \quad (30)$$

or

$$2d_{\text{TV}}(p, q) \tanh d_{\text{TV}}(p, q) \leq \frac{D(p\|q) + D(q\|p)}{2}. \quad (31)$$

Equivalently, as  $\text{atanh } x = \frac{1}{2} \ln \frac{1+x}{1-x}$ , we obtain

$$d_{\text{TV}}(p, q) \ln \frac{1 + d_{\text{TV}}(p, q)}{1 - d_{\text{TV}}(p, q)} \leq \frac{D(p\|q) + D(q\|p)}{2}. \quad (32)$$

In fact this bound is known to be tight, since equality is achieved for binary distributions  $p = (\epsilon, 1 - \epsilon)$  and  $q = (1 - \epsilon, \epsilon)$ , for any  $0 < \epsilon < 1$  [33]. This bound is the key ingredient for estimating the thermodynamic cost of a bit switch in the next section.

## 5. Application I: trade-off between reliability and thermodynamic cost for fast computation and other switched systems

### 5.1. A model for one-bit memories

We propose the following model for a one-bit memory. We set up a time-varying Markov process over a state space  $\Omega \ni \omega$  with two possible rates matrices  $L_p$  and  $L_q$ . When the rate matrix is  $L_p$ , then the Markov chain converges to the (unique) stationary distribution  $p$ , thereby encoding a ‘zero’, and converges to  $q$  when the rate matrix is  $L_q$ , thereby encoding ‘one’. An external device switches the rate matrix from  $L_p$  to  $L_q$  (thus ‘writing a one’ in the memory) or from  $L_q$  to  $L_p$  (thus ‘writing a zero’). Another device reading the memory must in fact solve an inference, or decoding problem: from the value of an observable on  $\Omega$  it must decide whether the Markov chain is more likely to encode a ‘zero’ (i.e. obey probability distribution  $p$ ) or a ‘one’ (i.e. obey probability distribution  $q$ ). If  $p$  and  $q$  happen to have disjoint supports in  $\Omega$  then the reading is errorless but usually  $p$  and  $q$  overlap, leaving the possibility of a wrong conclusion. We assume that the switch from  $L_p$  rates to  $L_q$  rates is instantaneous, or (more realistically) much faster than any other timescale of the Markov chain. Moreover, a reading, or a new writing operation occurs after the Markov process has converged to its stationary distribution (a reasonable assumption in practice). We defer to the end of the section a more concrete description of the specific physical implementations to which this theory applies.

### 5.2. Entropy production and error probability associated to a bit switch

In this context,  $\Delta\sigma_{\text{switch}} = k_B \frac{D(p\|q) + D(q\|p)}{2}$  has a thermodynamic interpretation: it is the non-adiabatic (also called Hatano–Sasa) entropy production associated to the relaxation process occurring in the transition between a ‘zero’ and a ‘one’, averaged over the two transitions. To show this we start from the splitting of the total entropy production rate  $\dot{\sigma}_{\text{total}} = \dot{\sigma}_{na} + \dot{\sigma}_{hk}$  into the non-adiabatic ( $\dot{\sigma}_{na} \geq 0$ ) and the adiabatic, or housekeeping ( $\dot{\sigma}_{hk} \geq 0$ ) contributions, the latter being zero only for detailed balanced dynamics. The non-adiabatic entropy production rate is defined as

$$\dot{\sigma}_{na} = -k_B \sum_{\omega} \dot{p}(t) \ln \frac{p(t)}{p_{\text{st}}(t)} = k_B \dot{H}(p(t)) + k_B \sum_{\omega} \dot{p}(t) \ln p_{\text{st}}(t), \quad (33)$$

where  $H(p(t)) = -\sum_{\omega \in \Omega} p(t) \ln p(t)$  is the Shannon entropy of  $p(\omega, t)$  the time-dependent solution of the master equation with rates  $L(t)$  switching between  $L_p$  and  $L_q$ . The stationary solution  $p_{\text{st}}(t)$  of the master equation is  $p$  (resp.  $q$ ) when  $L(t) = L_p$  (resp.  $L(t) = L_q$ ). Thus, the non-adiabatic entropy production associated to the transition from  $p$  to  $q$  is

$$\sigma_{na}(p \rightarrow q) = k_B \left[ H(q) - H(p) + \sum_{\omega} (q - p) \ln q \right]. \quad (34)$$

The non-adiabatic entropy production averaged over two consecutive switches reads

$$\Delta\sigma_{\text{switch}} = \frac{k_B}{2}(\sigma_{na}(p \rightarrow q) + \sigma_{na}(q \rightarrow p)) = k_B \frac{D(p\|q) + D(q\|p)}{2}, \quad (35)$$

where the opposite Shannon entropy variations cancel each other. This derivation holds even if  $p$  and  $q$  are non-equilibrium stationary distributions: in that case the term above counts the entropy production associated to the relaxation excluding the housekeeping (yet time-varying) entropy production needed to sustain the memory (i.e. to maintain the violation of detailed balance) and the instantaneous work to manipulate it (i.e. switch from  $L_p$  to  $L_q$  and vice versa). It is also true if  $p$  and  $q$  are quasi-stationary instead of truly stationary, i.e. stay approximately constant over the time scale in which reading and writing operations occur, as is the case for example in bistable ('flip-flop') memories. In case where  $p$  and  $q$  are equilibria at temperature  $T$ ,  $\Delta\sigma_{\text{switch}}T$  is the heat dissipated into the relaxation processes of the bit switching. This is the case in some recent experimental setups [34, 35]. Conversely,  $\Delta\sigma_{\text{switch}}$  is not directly relatable to the dissipated work in typical technological implementations in which the housekeeping entropy is non-zero (e.g. CMOS inverters or bistable memories [36, 37]). In case of isothermal transformations,  $\Delta\sigma_{\text{switch}}T$  is only a lower bound on the dissipation over the cycle, which is  $\Delta\sigma_{\text{total}}T$ .

The total variation distance  $d_{\text{TV}}(p, q)$  also has a relevant interpretation. Suppose we observe the full state  $\omega$  of the system (e.g. the charge in all capacitances) at a time instant  $t$  and we try to infer whether the system is in a 'zero' or a 'one', i.e. follows distribution  $p$  or  $q$ . If a 'zero' or a 'one' are *a priori* equally likely, then the optimal inference is the maximum likelihood estimator: if  $p(\omega) > q(\omega)$  then we infer a 'zero', otherwise we infer a 'one'. Thus the observable  $f = \text{sign}(p - q)$  is the optimal 'reading' of the bit. In this context,  $\epsilon = \frac{1-d_{\text{TV}}(p,q)}{2}$  is precisely the probability of error, i.e. of reading a 'zero' while the system obeys distribution  $q$  or a 'one' when it obeys  $p$ . Hence, the response intensity relation for the sign observable (32) can be rewritten as

$$k_B(1 - 2\epsilon) \ln(\epsilon^{-1} - 1) \leq \Delta\sigma_{\text{switch}} \leq \Delta\sigma_{\text{total}}, \quad (36)$$

where in the last equality we used that the total entropy production  $\Delta\sigma_{\text{total}}$  of the switch is obtained by adding the non-negative housekeeping entropy production. Equation (36) expresses a relation between error probability and thermodynamic cost of the computation. We remind that this relation—the main result of this section—is deduced under the assumption of a Markov process undergoing an abrupt (instant) switch between the two dynamics.

### 5.3. Reliable vs noisy computations

In the limit of reliable computations,  $\epsilon \ll 1$ , (36) simplifies to  $k_B \ln \epsilon^{-1} \leq \Delta\sigma_{\text{switch}}$ , or equivalently

$$\epsilon \geq e^{-\Delta\sigma_{\text{switch}}/k_B}. \quad (37)$$

A similar relation has been obtained under particular technological assumptions fitting the current technological paradigm (e.g. circuit with linear capacitances operating under Gaussian thermal noise) [38].

In the opposite noisy computation limit,  $d_{\text{TV}} = 1 - 2\epsilon \ll 1$  (i.e.  $\epsilon \approx 1/2$ ), (36) simplifies to

$$2k_B d_{\text{TV}}^2 \leq \Delta\sigma_{\text{switch}}. \quad (38)$$

Note that in this limit, the output has very low information in the sense of Shannon, i.e. the mutual information between writing (alternating between  $L_p$  or  $L_q$  with probability 1/2) and the reading is very low (with error probability close to 1/2). As a matter of fact, the mutual information is equal (in nats) to

$$I = \ln 2 - [\epsilon \ln \epsilon^{-1} + (1 - \epsilon) \ln (1 - \epsilon)^{-1}] \approx d_{\text{TV}}^2 / 2, \quad (39)$$

where the first term is the Shannon entropy of the writing process and the second term in square brackets is the conditional entropy of the writing given output of the reading process. Therefore, the cost of switching a bit in this context is

$$4k_B I \leq \Delta\sigma_{\text{switch}} \leq \Delta\sigma_{\text{total}}. \quad (40)$$

To speak in terms of energy, the cost of switching a bit abruptly in an environment of temperature  $T$  is at least  $4k_B T$  per nat, or  $2.8k_B T$  per Shannon bit, independently of the (low) reliability. This is in contrast with the low noise limit (37), where the cost per bit switch grows unboundedly as we approach perfect reliability. This supports quantitatively the common knowledge that noisy circuits may be more energetically efficient than low-noise circuits. Of course, managing noisy information may require carefully designed noise-tolerant algorithms [36, 39, 40].

#### 5.4. Discussion on the energy-error-speed trade-off

To make things more concrete, we consider the example of the NOT gate (also known as the inverter), an elementary electronic circuit ubiquitous in digital computing devices, as it is a building block for more complex circuits. This circuit is controlled by an external voltage  $u(t)$  which takes the value  $u_{\min}$  or  $u_{\max}$ . This voltage determines the random dynamics of the charges flowing into the capacitances of the inverter. When stationarity is reached, an output capacitance has a *mean* voltage close to  $u_{\max}$  (encoding a ‘one’) when the input voltage is  $u_{\min}$  and conversely. The reading of the bit is in practice a thresholding of the output capacitance voltage: if it takes a value larger than  $(u_{\min} + u_{\max})/2$  we consider it encodes a ‘one’, otherwise a ‘zero’. Due to random fluctuations however, an erroneous reading may occur. The technological details do not matter here: it suffices to know that the dynamics of such a circuit is well modeled indeed by a Markov process. We refer the reader to [36, 37] for explicit and realistic constructions of Markov chains for inverters and more complex memories.

We made the assumption of an abrupt, instant switching between two rate matrices. In the practice of a digital circuit such as an inverter, this would amount to an instant switch between voltages  $u_{\min}$  and  $u_{\max}$ . Since these voltages are themselves the output of another circuit, they do not evolve instantly but with a certain time constant. Taking these time scales into account would need to correct the analysis above. Nevertheless, these time constants tend to be made as short as possible in current digital technologies, and the abrupt switching can be seen as an ideal limiting case.

Let us remind that the actual reading of the bit in practice is not always given by the maximum-likelihood binary observable  $\text{sign}(p - q)$  as we have assumed here, but for instance by  $\text{sign}(f)$ , where  $f$  is a physical observable such as the voltage on one specific capacitance. If the two differ, the reading process may be suboptimal, resulting in an extra error probability. From Chebyshev–Cantelli’s inequality alone, assuming  $f$  centered and  $\text{Var}_p f = \text{Var}_q f$  for simplicity, we know that an inference based on  $\text{sign}(f)$  will have an error  $\epsilon \leq (1 + R_{p,q}(f)/4)^{-1}$ , which is a very slow decrease. If more is known about the distribution of  $f$ , for instance if it is Gaussian, then we may derive an exponentially fast decrease of the error with  $R_{p,q}(f)$ , as indicated from Chernov’s bound for instance, or bounds in [41].

Remember that the main result of this section, (36), is derived from (27) applied to the observable  $\text{sign}(p - q)$  (which is the outcome of the bit reading process). If our goal is to obtain a lower bound on  $\sigma_{\text{switch}}$  that is as tight as possible, one should rather use the response intensity of the maximum-response-intensity observable  $m = 2(p - q)/(p + q)$  or an observable close to it. In the high reliability limit ( $\epsilon \ll 1$ ), the distributions  $p$  and  $q$  have a small overlap, and the binary observable  $\text{sign}(p - q)$  is always close to  $m$  indeed. In the low accuracy limit however, the two might differ substantially, and the response of other observables may offer a better lower bound on entropy production. Consequently, the bound (40) may be loose in these cases.

Importantly, note that the bound (40) is unrelated to the similar-looking Landauer's bound. Landauer's bound concerns the *erasure* of a bit, and is relevant universally. We know, in agreement with Landauer's bound, that switching a bit from zero to one and conversely can in principle be done at zero dissipation, with underdamped time-varying circuits or with infinitely slow overdamped circuits. We place ourselves at another end of the spectrum, with abrupt (as fast as possible) switching in a circuit undergoing a relaxation to stationarity. A full theory encompassing the energy-accuracy trade-off in all the spectrum of speeds is beyond the scope of this paper. Our results here are, as far as we know, among the first towards a trade-off for general classes of systems at non-zero speeds. See however [42] for energy-optimal erasure in finite time on the real line.

Finally, even though we discussed the case of a bit switch in this section, note that dynamical systems switching abruptly between two or more regimes are common in physics and engineering beyond the realm of computation, and the results discussed here can in principle be adapted for those (see appendix D). More abstractly, every inference problem involving a binary decision can benefit from the same analysis. For example, in [22] it was considered the problem of deciding whether a given trajectory of a stationary Markov chain is unfolded forward or backward in time. Equation (9) in [22] coincides with our (36) with  $\Delta\sigma_{\text{switch}}$  being the total entropy production of the process. Adapting our derivation to this context (see appendix E), we can indeed extend the validity of the result of [22], showing that it holds under very general circumstances, e.g. independently of the choice of a certain stopping criterion and for non-Markovian dynamics.

## 6. Application II: classical speed limits

### 6.1. What is a classical speed limit?

Classical speed limits (SLs) are lower bounds on the time needed for a driven system to reach a target state from an initial state, in terms of thermodynamic or kinetic quantities. The main such result [18, 19] states that if a non-autonomous finite-state Markov chain starts from state probability  $p(0) = p$  and finishes in state probability  $p(\tau) = q$ , then the time  $\tau$  to do so is lower bounded as follows:

$$\tau \geq \frac{2d_{\text{TV}}^2(p, q)}{\Delta\sigma\bar{A}}. \quad (41)$$

In this inequality,  $\Delta\sigma$  is the (non-adiabatic, or Hatano–Sasa) entropy production along the trajectory that is associated with the relaxation to the instantaneous stationary state. The latter is defined at each time  $t$  as the probability distribution that would be reached by the system if the driving stopped, i.e. the transition rate became time-independent for all times larger than  $t$ . The activity  $\bar{A}$  is the time-average activity rate, i.e. expected number of jumps per time unit. In (41), a thermodynamic term (non-adiabatic entropy production) quantifies how far we stand

from the stationary distribution, and a kinetic term (activity) quantifies how fast the system evolves.

In this section we adopt the view that a classical SL is any inequality that can be written in the form

$$\tau \geq h(\text{Dist}, \text{Therm}, \text{Kin}) \quad (42)$$

where Dist is the distance between  $p$  and  $q$ , for any notion of distance between probability distributions, Therm is a thermodynamic quantity capturing how far  $p(t)$  is from equilibrium or stationarity, and Kin is a kinetic quantity which measures the ‘fastness’ of the dynamics. If the time  $\tau$  and the distance are very small (infinitesimal), we can speak of *local classical SL*. We leverage the concepts encountered above, such as the maximum response intensity between  $p$  and  $q$ , to derive new classical SLs involving various statistical distances, thermodynamic and kinetic quantities of interest.

### 6.2. Bounding Le Cam’s distance with an integral along the trajectory

Le Cam’s distance  $d_{\text{LeCam}}(p, q)$ , via (17), captures the maximum response intensity between  $p$  and  $q$ . The key observation is that the triangle inequality allows to write

$$d_{\text{LeCam}}(p, q) \leq \sum_{k=0}^{K-1} d_{\text{LeCam}}(p(k\Delta t), p((k+1)\Delta t)) \quad (43)$$

for  $\tau = K\Delta t$ , where  $\Delta t$  can then be taken arbitrarily small, and  $K$  very large. In the limit of very small  $\Delta t$ , Le Cam’s distance essentially coincides with Fisher information metric:

$$d_{\text{LeCam}}^2(p, p + \Delta p) = \frac{1}{4} \sum_{\omega} \frac{\Delta p^2}{p} + \mathcal{O}(\Delta p^3). \quad (44)$$

From the triangle inequality and the local behavior of the distance, we find:

$$d_{\text{LeCam}}(p, q) \leq \frac{1}{2} \int_0^\tau \left\| \frac{\dot{p}}{\sqrt{p}} \right\|_2 dt \quad (45)$$

where  $\|\cdot\|_2$  is the usual Euclidean norm (square root of sum of squares). To go further, we need to compute and bound  $\dot{p}$ , as we do in next sections.

### 6.3. Forward and backward master equation

For Markov chains the evolution of  $p(t)$  is dictated by a transition rate matrix  $L(t)$ , to which an instantaneous stationary distribution  $p_{\text{st}}(t)$  is associated. We thus have the master equation

$$\dot{p}(t) = p(t)L(t), \quad (46)$$

as commonly written in Markov chain literature notation, i.e. with the state probabilities as a row vector, and the matrix elements  $L_{\omega\omega'} (\omega \neq \omega')$  denoting the transition rate from state  $\omega$  to state  $\omega'$ . The diagonal entry is  $L_{\omega\omega} = -\sum_{\omega' \neq \omega} L_{\omega\omega'}$ , so that each row of the matrix  $L$  sums to zero. We assume for simplicity that at each time instant,  $L(t)$  admits a unique stationary probability distribution  $p_{\text{st}}(t)$ , defined as the left eigenvector satisfying  $p_{\text{st}}(t)L(t) = 0$ .

As it will be useful later, we also write the master equation backward in time,

$$-\dot{p}(t) = p(t)L^*(t) \quad (47)$$

where the transition rate matrix is  $L^* = P^{-1}L^TP - P^{-1}\dot{P}$ , and  $P = \text{diag}(p)$  is the diagonal matrix created from  $p$ . This is proved as follows: the off-diagonal entries of  $PL$  and  $(PL^*)^T$  coincide, because the entry  $\omega\omega'$  of  $PL$  represents the probability flow from state  $\omega$  to state  $\omega'$ , and the entry  $\omega'\omega$  of  $PL^*$  represents the probability flow from state  $\omega'$  to state  $\omega$  in the backward dynamics, which is of course the same. The diagonal entries of  $L^*$  must be such that each row of  $L^*$  sums to zero. The diagonal entries of  $L$  represent the *escape rates* from each state, while the diagonal entries of  $L^*$  represent the escape rates in the backward dynamics, thus the *entrance rates* in the forward dynamics [43].

#### 6.4. Local speed limit

In order to estimate (45), we want to find an upper bound on  $\|\dot{p}P^{-1/2}\|_2$ , knowing that  $\dot{p} = pL$ . This will lead to what we can call a local classical SL.

Knowing that  $p_{\text{st}}L = 0$  (expressing that  $p_{\text{st}}$  is a stationary state), we can write  $\dot{p} = (p - \alpha p_{\text{st}})L$ , for any scalar coefficient  $\alpha$  of our choice. We can also write:

$$\dot{p}P^{-1/2} = (p - \alpha p_{\text{st}})LP^{-1/2} \quad (48)$$

$$= (\sqrt{p} - \alpha p_{\text{st}}P^{-1/2})P^{1/2}LP^{-1/2}. \quad (49)$$

It is useful to introduce the two-norm of a square matrix  $A$ , defined as the maximum ratio  $\frac{\|xA\|_2}{\|x\|_2}$  over all non-zero (row) vectors  $x$ . It turns out to be equal to the maximum singular vector of  $A$ , i.e. the square root of the spectral radius  $\rho$  (eigenvalue of largest magnitude) of the symmetric matrix  $AA^T$ . This is denoted

$$\|A\|_2 = \sqrt{\rho(AA^T)} = \sqrt{\rho(SAA^TS^{-1})}, \quad (50)$$

for any invertible square matrix  $S$  (encoding a linear change of coordinates, which does not affect eigenvalues). Thus we obtain from (49):

$$\|\dot{p}P^{-1/2}\|_2 \leq \|\sqrt{p} - \alpha p_{\text{st}}P^{-1/2}\|_2 \|P^{1/2}LP^{-1/2}\|_2. \quad (51)$$

First, we estimate  $\|P^{1/2}LP^{-1/2}\|_2$  as

$$\|P^{1/2}LP^{-1/2}\|_2 = \sqrt{\rho(P^{1/2}LP^{-1/2}P^{-1/2}L^TP^{1/2})} \quad (52)$$

$$= \sqrt{\rho(LP^{-1}L^TP)} \quad (53)$$

$$\leq \sqrt{\|L\|_\infty \|P^{-1}L^TP\|_\infty}. \quad (54)$$

Here we have used the matrix norm  $\|A\|_\infty$ , which is an upper bound on the spectral radius ( $\rho(A) \leq \|A\|_\infty$ ) and submultiplicative ( $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$ ). It is defined as the maximum one-norm of a row of  $A$ , or equivalently as  $\max \frac{\|xA\|_1}{\|x\|_1}$ . Thus  $\|L\|_\infty$  is twice the maximum escape rate from each state, which we denote  $A_{\text{max.escape}} = \max_\omega L_{\omega\omega}$ . The one-norm of  $P^{-1}L^TP$  is obtained by noticing that the off-diagonal entries of this matrix are the same as  $L^*$  (the matrix

of backward rates), while the diagonal entries are the same as  $L$ . Thus the one-norm of  $P^{-1}L^TP$ , denoted  $A_{\max, \text{entrance}}$ , is the largest sum of escape rate and entrance rate (which is the escape rate in the backward random walk) on a state. Therefore, we arrive at

$$\|P^{1/2}LP^{-1/2}\|_2 \leq \sqrt{\|L\|_\infty \|P^{-1}L^TP\|_\infty} \quad (55)$$

$$\leq \sqrt{2A_{\max, \text{escape}}(A_{\max, \text{exit}} + A_{\max, \text{entrance}})} \quad (56)$$

$$\leq 2 \max(A_{\max, \text{escape}}, A_{\max, \text{entrance}}) \quad (57)$$

$$= 2A_{\max}, \quad (58)$$

where  $A_{\max} = \max(A_{\max, \text{escape}}, A_{\max, \text{entrance}})$  is the *maximum activity rate*. Second, we estimate  $\|\sqrt{p} - \alpha p_{\text{st}}P^{-1/2}\|_2$ :

$$\|\sqrt{p} - \alpha p_{\text{st}}P^{-1/2}\|_2^2 = \sum_{\omega} [p(\omega) - 2\alpha p_{\text{st}}(\omega) + \alpha^2 p_{\text{st}}^2(\omega)/p(\omega)] \quad (59)$$

$$= 1 - 2\alpha + \alpha^2 \sum_{\omega} p_{\text{st}}^2(\omega)/p(\omega) \quad (60)$$

$$= 1 - 2\alpha + (1 + \chi^2(p_{\text{st}}\|p))\alpha^2. \quad (61)$$

Choosing  $\alpha = 1/(1 + \chi^2(p_{\text{st}}\|p))$ , which minimizes (61), we obtain:

$$\|\sqrt{p} - \alpha p_{\text{st}}P^{-1/2}\|_2^2 = \frac{\chi^2(p_{\text{st}}\|p)}{1 + \chi^2(p_{\text{st}}\|p)} \quad (62)$$

$$\leq \min(1, \chi^2(p_{\text{st}}\|p)). \quad (63)$$

With these estimates, our bound (51) turns into the local classical SL

$$\|\dot{p}/\sqrt{p}\|_2 \leq 2 \sqrt{\frac{\chi^2(p_{\text{st}}\|p)}{1 + \chi^2(p_{\text{st}}\|p)}} A_{\max}. \quad (64)$$

This is indeed a classical SL in the form (42) as an upper bound on the speed  $\dot{p}$  is also a lower bound on the (short) time  $\Delta t$  needed to achieve a certain (short) displacement  $\Delta p = \dot{p}\Delta t + \mathcal{O}(\Delta t^2)$ . Referring to (42), our thermodynamic quantity is  $\chi^2(p_{\text{st}}\|p)$ , which quantifies, as we have seen, the maximum response intensity to the change from  $p$  to stationarity  $p_{\text{st}}$ , thus capturing in a meaningful way whether  $p$  is far from stationarity. Our kinetic quantity is  $A_{\max}$ , which bounds the entrance or escape at any state. In other words,  $1/A_{\max}$  can be interpreted as the fastest time scale at any state in any direction of time.

### 6.5. The classical SL for Le Cam's distance

Integrating the local bound (64) into (45), we obtain a SL for the time interval  $[0, \tau]$  between  $p = p(0)$  and  $q = p(\tau)$ :

$$d_{\text{LeCam}}(p, q) \leq \int_0^\tau \sqrt{\frac{\chi^2(p_{\text{st}}\|p)}{1 + \chi^2(p_{\text{st}}\|p)}} A_{\max} dt. \quad (65)$$

In order to obtain separate thermodynamic and kinetic factors, we can for instance bound the integral as follows:

$$\left| \int_0^\tau f g \, dt \right| \leq \sup_t |f| \int_0^\tau |g| \, dt. \quad (66)$$

In this way we get a novel classical SL:

$$d_{\text{LeCam}}(p, q) \leq \sup_{0 \leq t \leq \tau} \sqrt{\frac{\chi^2(p_{\text{st}} \| p)}{1 + \chi^2(p_{\text{st}} \| p)}} \bar{A}_{\max} \tau \quad (67)$$

where  $\bar{A}_{\max} = \tau^{-1} \int_0^\tau A_{\max}(t) \, dt$  is the time-average of the maximum activity. This evidently provides a lower bound on  $\tau$ , in terms of total distance, thermodynamic and kinetic factors, following the template (42). Note that since the thermodynamic term is bounded by 1, we have the purely kinetic SL:

$$d_{\text{LeCam}}(p, q) \leq \bar{A}_{\max} \tau. \quad (68)$$

This bound is acceptable far from stationarity, when  $\chi^2(p_{\text{st}} \| p) \gg 1$  for some part at least of the trajectory. Nevertheless, the thermodynamic factor is useful when driving close to stationarity, when  $\chi^2(p_{\text{st}} \| p) \ll 1$  (low asymmetric response intensity), yielding a much tighter bound than (68).

### 6.6. A classical SL with the total variation distance

We now apply the same methodology as for Le Cam's distance to the total variation distance. Locally, we find:

$$d_{\text{TV}}(p(t), p(t + \Delta t)) = \frac{1}{2} \|\Delta p\|_1 = \frac{1}{2} \|\dot{p}\|_1 \Delta t + \mathcal{O}(\Delta t^2). \quad (69)$$

Thus, from the triangle inequality we obtain

$$d_{\text{TV}}(p, q) \leq \frac{1}{2} \int_0^\tau \|\dot{p}(t)\|_1 \, dt,$$

in which the integrand in the rhs can be estimated as

$$\frac{1}{2} \|\dot{p}(t)\|_1 = \frac{1}{2} \|pL\|_1 \quad (70)$$

$$= \frac{1}{2} \|(p - p_{\text{st}})L\|_1 \quad (71)$$

$$\leq \frac{1}{2} \|(p - p_{\text{st}})\|_1 \|L\|_\infty \quad (72)$$

$$\leq \frac{1}{2} \|(p - p_{\text{st}})\|_1 2 \max_\omega |L_{\omega\omega}| \quad (73)$$

$$= 2d_{\text{TV}}(p, p_{\text{st}}) A_{\max, \text{exit}}, \quad (74)$$

yielding

$$d_{\text{TV}}(p, q) \leq 2 \int_0^\tau d_{\text{TV}}(p, p_{\text{st}}) A_{\text{max,escape}} dt. \quad (75)$$

Using (66), the latter can be further bounded as

$$d_{\text{TV}}(p, q) \leq 2 \sup_{0 \leq t \leq \tau} d_{\text{TV}}(p, p_{\text{st}}) \bar{A}_{\text{max,escape}} \tau, \quad (76)$$

which constitutes an example of classical SL involving the total variation distance both as the distance between initial and final state, and as a thermodynamic measure of distance to stationarity. It can be related to available free energy  $D(p \| p_{\text{st}})$  through Pinsker's inequality

$$d_{\text{TV}}(p, p_{\text{st}}) \leq \sqrt{\frac{1}{2} D(p \| p_{\text{st}})}, \quad (77)$$

thus giving the classical SL

$$d_{\text{TV}}(p, q) \leq \sup_{0 \leq t \leq \tau} \sqrt{2 D(p \| p_{\text{st}})} \bar{A}_{\text{max,escape}} \tau. \quad (78)$$

It resembles the original classical SL obtained in [18, 19], although they are not directly comparable. The main conceptual difference is that here our thermodynamic factor is the total available free energy (in case of detailed balanced dynamics), rather than its dissipation rate.

Note that this can be improved by using other bounds on total variation distance than Pinsker's, for instance Bretagnolle–Huber's bound [44]

$$d_{\text{TV}}(p, p_{\text{st}}) \leq \sqrt{1 - e^{-D(p \| p_{\text{st}})}}, \quad (79)$$

which is worse than Pinsker's close to stationarity but better far from stationarity. We thus find another classical SL:

$$d_{\text{TV}}(p, q) \leq 2 \sup_{0 \leq t \leq \tau} \sqrt{1 - e^{-D(p \| p_{\text{st}})}} \bar{A}_{\text{max,escape}} \tau. \quad (80)$$

From this we have again a purely kinetic SL:

$$d_{\text{TV}}(p, q) \leq 2 \bar{A}_{\text{max,escape}} \tau. \quad (81)$$

This SL is useful if  $p$  is far from stationarity (total variation distance close to one), but extremely loose when driving the distribution in a quasi-stationary fashion. An analogous fact is true for precision: kinetic uncertainty relations [30] are tighter than thermodynamic ones for large entropy production.

Note that one can easily obtain many variants of the SLs, following the same methodology with different bounds or distances. A few ones are highlighted in appendix F.

## 7. Application III: simplicity implies robustness, complexity implies fragility

In this section we go back to the problem of bounding the nonlinear response of a physical system. Given a reference distribution  $q$  on a space  $\Omega$ , and a perturbed distribution  $p$ , the maximum of (asymmetric) response intensity over all observables on  $\Omega$  is  $\chi^2(p \| q) = \sum_{\omega} p^2 / q - 1$ , as derived in section 4.3. We now bound this further, for a given unperturbed distribution  $q$ , over a range of possible perturbations  $p$ , and find in two different situations that only 'complex'  $q$  allow large maximal perturbations  $\chi^2(p \| q)$ .

### 7.1. In trajectory space: Langevin dynamics

First, we focus on  $\Omega$  being the space of trajectories of a stochastic system and we rewrite the probability density  $p(\omega) = e^{\mathcal{A}_p(\omega)}$  (resp.  $q(\omega) = e^{\mathcal{A}_q(\omega)}$ ) as the exponential of an action functional. To be concrete, we consider the case of diffusive dynamics given (in the unperturbed case  $q$ ) by the Langevin equation

$$\dot{x}(t) = \mu F(x(t)) + \sqrt{2D}\xi(t) \quad (82)$$

for the vector  $x \in \mathbb{R}^{dN}$  describing, e.g.  $N$  particles in  $d$  dimensions. Equation (82) features a delta-correlated Gaussian noise  $\xi$ , a positive definite symmetric diffusion matrix  $D$  and mobility matrix  $\mu$  (not necessarily connected by the Einstein relation  $D = k_B T \mu$ ). It corresponds to the action

$$\mathcal{A}_q(\omega) = -\frac{1}{4} \int_0^\tau dt (\dot{x}(t) - \mu F(x(t))) D^{-1} (\dot{x}(t) - \mu F(x(t)))^T \quad (83)$$

for the trajectories  $\omega = \{x(t) : t \in [0, \tau]\}$ , when the Ito convention is used. If  $p$  and  $q$  are absolutely continuous, i.e. the ratio  $p/q$  (more rigorously, the Radon–Nikodym derivative) exists and is finite, the computation of the maximal response reduces to an exponential average over unperturbed trajectories

$$\chi^2(p\|q) = \langle e^{2(\mathcal{A}_p - \mathcal{A}_q)} \rangle_q - 1, \quad (84)$$

that might be much less involved than an average w.r.t. the perturbed  $p$  ( $q$  can refer to an equilibrium or non-interacting system, for instance). Perturbations that satisfy (84) are, for example, changes of the drift, such as  $F(x) \rightarrow F(x) + g(x)$ . Instead, a perturbed probability density  $p$  generated by a change in the matrix  $D$  is not absolutely continuous w.r.t.  $q$  (and response in trajectory space is known to require explicit regularization procedures [45, 46]). Restricting us to a change in the external forces  $F(x) \rightarrow F(x) + g(x)$ , we obtain

$$\chi^2(p\|q) = \left\langle e^{\int_0^\tau dt [\mu g D^{-1}(\dot{x} - \mu F) - \frac{1}{2} \mu g D^{-1}(\mu g)^T]} \right\rangle_q - 1. \quad (85)$$

In general (85) is hard to compute as it involves averaging a functional of the trajectories and thus requires knowledge of the time propagator. Possibly, a saddle point calculation can be performed in some large-deviation regimes, e.g. long time  $\tau \rightarrow \infty$ , weak noise  $D \rightarrow 0$  (entry-wise), and in some proper thermodynamic limit involving  $N \rightarrow \infty$ . For small perturbations, we retrieve Dechant–Sasa’s bound on the linear response (see (12)),

$$\max_f R_{p\|q}(f) = \frac{1}{2} \int_0^\tau dt \langle \mu g D^{-1}(\mu g)^T \rangle_q. \quad (86)$$

Here we point out how the complexity of the system (spread in space and range of interactions) potentially leads to large responses in paradigmatic models of energy and mass transport. To show that we consider as the unperturbed system independent particles in a stationary state, subject to a single-particle force  $F(x) = (F_1(x_1), \dots, F_N(x_N))^T$  and in contact with separate thermal baths at temperature  $T_i$ . Hence, we set  $\mu$  equal to the identity matrix (choosing appropriate time units) and  $D$  diagonal with elements  $D_{ii} = k_B T_i \mu_i$ . The perturbation  $g(x) = (g_1(x), \dots, g_N(x))$  may consist of an arbitrary interaction force coupling different

particles. The maximal linear response is thus given by

$$\frac{1}{\tau} \max_f R_{p\|q}(f) = \sum_{i=1}^N \mu_i \frac{\langle \|g_i\|_2^2 \rangle_q}{2k_B T_i} \leq M^2 \kappa^2 \sum_{i=1}^N \frac{\langle \|x_i\|_2^2 \rangle_q}{2k_B T_i}, \quad (87)$$

where the second equality holds by assuming that the interactions can be modeled to leading order as a linear force of strength  $\kappa$  involving  $M$  neighbors, and by taking  $x = (x_1, \dots, x_N)^T$  as the displacement from the particle rest positions (i.e.  $\langle x \rangle_q = 0$ ). This general setting covers standard models of heat conducting lattices and loaded molecular motors (see e.g. [47, 48]). In the first case,  $F(x)$  is an arbitrary nonlinear potential force; in the second  $F(x)$  is typically a tilted sinusoidal potential, the dynamics is restricted by spatial periodic boundary conditions within a period, and  $T_i = T$  for all  $i$ . Note that for the observables of interest, such as the time-integrated heat and particle currents, the scaling by  $1/\tau$  ensures that the lhs of (87) is independent of  $\tau$ . In both cases the more ‘complex’ is the system (larger spreading in space  $\langle \|x_i\|_2^2 \rangle_q$ ) and the perturbation (larger range of interactions  $M$ ), the larger the maximal linear response.

## 7.2. In finite state spaces

Second, we take  $\Omega$  to be the state space of a stochastic system. For a finite space  $\Omega$ , we now consider the maximum  $\chi^2(p\|q)$  over all possible perturbations  $p$ . The perturbation leading to highest possible response is the distribution  $p$  concentrated on  $\omega$  such that  $q(\omega) = q_{\min}$  is minimal. Thus

$$\max_p \chi^2(p\|q) = \frac{1}{q_{\min}} - 1. \quad (88)$$

We take  $1/q_{\min}$  as an index of complexity of the system. If  $\Omega$  has cardinality  $N$  then  $1/q_{\min} \geq N$ . Inversely, if  $q$  is relatively uniform over a few states, then  $1/q_{\min}$  is small. This is quantified by the following inequalities:

$$e^{H(q)} \leq N \leq \frac{1}{q_{\min}} \leq e^{H(q)} \frac{q_{\max}}{q_{\min}} \leq N \frac{q_{\max}}{q_{\min}}, \quad (89)$$

where  $H(q) = \sum_{\omega \in \Omega} q \ln \frac{1}{q}$  is the Shannon entropy, with bounds

$$\frac{1}{q_{\max}} \leq e^{H(q)} \leq N \leq \frac{1}{q_{\min}}. \quad (90)$$

We can always write  $q(\omega)$  in an exponential form  $q(\omega) = e^{-\beta(E(\omega)-F)}$ , where  $F = \langle E \rangle - \beta^{-1}H(q)$ . This is especially relevant when  $q$  is an equilibrium distribution, where  $E$  is an energy,  $\beta^{-1} = k_B T$  and  $F$  the equilibrium free energy. In this case we can write:

$$\frac{1}{q_{\min}} = e^{-\beta E_{\max} + F} \leq e^{H(q)} e^{\beta(E_{\max} - E_{\min})}. \quad (91)$$

Thus as a general rule we find that ‘simple’ systems (with low  $1/q_{\min}$ ) are ‘robust’, in that the response of any observable to any perturbation does not go beyond a few standard deviations. Conversely, ‘complex’ systems (e.g. with a large entropy or large energy spreading) are fragile, in that some observables are highly sensitive to some perturbations. This is reminiscent, although in another context, of the ‘complexity implies fragility’ principle advocated in

[49, 50]. If we see the response as a useful signal rather than a disturbance, then this means that simple systems are unable to transmit with a high signal-to-noise ratio.

This allows some insight on the value of simple models. Consider that typically nonequilibrium toy models have few states, e.g.  $N = 3$  for a minimal Markov chain able to sustain a current (without recurring to multiple transitions between two states). A typical example is offered by the enzyme kinetics describing substrate inhibition as a Markov chain on the three different states of the enzyme (free, and bounded to one or two substrate molecules) [51, 52]. If we mean to describe a biophysical process in this way, then typically the energy spread of the unperturbed system can be taken to be  $E_{\max} - E_{\min} \simeq k_B T$ . Therefore, we find  $|\langle f \rangle_p - \langle f \rangle_q| \lesssim \sqrt{3e^1 - 1} \sqrt{\text{Var}_q(f)} \approx 2.67 \sqrt{\text{Var}_q(f)}$ , namely, the response never exceeds 3 standard deviations irrespective of the perturbation strength. Thus, few-state models are by construction robust. Also, if results of an experiment were rationalized with a few-state model, they would be unable to tell whether the measured system is in or out of equilibrium.

### 7.3. Discussion

In these two situations, we propose a notion of ‘complexity’ related to the spreading of the state distribution of the system onto a large volume (variance of the state in the Langevin dynamics,  $1/q_{\min}$  for the finite state space case). We observe that ‘simple’ systems are bound to be robust, i.e. exhibit low response to perturbations. Complex systems on the other hand allow possibly large responses. It should be noted however that such a large response can only be reached by certain well-chosen observables, that depend on the global state of the system. For example consider a ‘local’ observable  $f$  that would only depend on a few degree of freedoms, e.g. the energy of a single particle, then it is expected that the formula (87) becomes asymptotically trivial for large  $N$ . This same phenomenon is also visible in the second situation (finite state space): in there nothing is said about the dynamics, which could be the non-Markovian coarse-graining of a large state space onto a few states only. Thus if the observable of interest can be expressed in terms of a few coarse-grained states, then the same bounds apply on the coarse-grained state space as well, where the complexity (here measured as  $1/q_{\min}$ ), thus the maximal response, can only be lower. Finally, let us observe that the dynamics on the small reduced (e.g. coarse-grained) dynamics is in general not known, and finding the tightest bounds for local observables is an open challenge [53].

## 8. Conclusions

We have explored bounds on the response of observables subject to a perturbation of the underlying probability distribution, introducing the response intensity as an adimensional ratio comparing the size of the response to the size of typical fluctuations. It turns out that bounds on the response intensity can be found by the Hilbert uncertainty relation, a recent generalization of TURs. These bounds allow to derive several nontrivial inequalities between the various distances or divergences that can be defined between probability distributions, some of them well-known and some of them novel. We believe that the novel SLs we have subsequently derived and the novel applications we have provided on switched systems (e.g. computing devices) and biophysical systems (e.g. the significance of simple models), together with other recent results of the stochastic thermodynamics literature, are only the tip of the iceberg of an emerging theory that characterizes the necessary relations between kinetics, thermodynamics and reliability of stochastic systems.

## Acknowledgments

J-CD and ME acknowledge funding from the Project INTER/FNRS/20/15074473 ‘TheCirco’ on Thermodynamics of Circuits for Computation, funded by the F.R.S.-FNRS (Belgium) and FNR (Luxembourg).

## Appendix A. Lower bounds on Le Cam’s distance

On the one hand, we derive from (10) lower bounds on Le Cam’s distance:

$$d_{\text{LeCam}}^2(p, q) \geq e^{D(p \parallel \frac{p+q}{2})} - 1 \geq D\left(p \parallel \frac{p+q}{2}\right) \quad (\text{A.1})$$

and

$$d_{\text{LeCam}}^2(p, q) \geq \frac{1}{2} e^{D(p \parallel \frac{p+q}{2})} + \frac{1}{2} e^{D(q \parallel \frac{p+q}{2})} - 1 \quad (\text{A.2})$$

$$\geq e^{\frac{1}{2}(D(q \parallel \frac{p+q}{2}) + D(p \parallel \frac{p+q}{2}))} - 1 \quad (\text{A.3})$$

$$\geq \frac{1}{2} \left( D\left(p \parallel \frac{p+q}{2}\right) + D\left(q \parallel \frac{p+q}{2}\right) \right) \quad (\text{A.4})$$

$$= d_{\text{JS}}^2(p, q). \quad (\text{A.5})$$

The latest quantity,  $d_{\text{JS}}(p, q)$ , defined as the square root of  $\frac{1}{2}(D(p \parallel \frac{p+q}{2}) + D(q \parallel \frac{p+q}{2}))$ , is also a distance (satisfying the triangle inequality), called the *Jensen–Shannon distance*.

## Appendix B. Equivalence of symmetric response intensity bounds with the TURs for antisymmetric observables

TURs were initially devised for  $\Omega$  being the set of trajectories of a (stochastic) physical system, and  $\mathcal{F}$  being the set of observables anti-symmetric under time-reversal (and possibly other constraints). In a more abstract version, regardless of any thermodynamic context, one considers an abstract  $\Omega$  endowed with an involution  $s : \Omega \rightarrow \Omega$  (i.e. a bijection whose square is the identity), which is not necessarily time-reversal but could be space reversal, spin reversal, etc [14]. The relations obtained as upper bounds on symmetric nonlinear response, such as (28) or (26), generalize results obtained previously in this literature, for  $p$  being the probability distribution on  $\Omega$ , and  $q$  being the symmetric distribution  $q(\omega) = p(s\omega)$ .

Conversely, these known results offer an alternative way to rederive the nonlinear response bounds, thanks to the following trick. Two distributions  $p$  and  $q$  on one same space  $\Omega$  can be recoded into a single distribution on the space  $\Omega \oplus \Omega$  (the disjoint union of two copies of  $\Omega$ ), with probability  $p/2$  on one copy of  $\Omega$  and  $q/2$  on the other copy. The involution  $s : \Omega \oplus \Omega \rightarrow \Omega \oplus \Omega$  simply swaps the two copies of  $\Omega$ . The observables  $f$  on  $\Omega$  can be recoded into antisymmetric observables  $f \oplus (-f)$  on  $\Omega \oplus \Omega$ . Applying the various bounds derived in [14] on antisymmetric observables we recover the bounds (26) or (28).

In conclusion there is an equivalence between the symmetric response intensity relation (28) and the TUR for antisymmetric observables with respect to arbitrary involutions.

### Appendix C. Symmetric Kullback–Leibler divergence and Jensen–Shannon distance

The bounds on Le Cam’s distance allow to compare the Jensen–Shannon (square) distance

$$d_{\text{JS}}^2(p, q) = \frac{D(p \parallel \frac{p+q}{2}) + D(q \parallel \frac{p+q}{2})}{2} \quad (\text{C.1})$$

with the symmetric Kullback–Leibler divergence, also called the Jeffreys divergence:

$$D(p, q) = \frac{D(p \parallel q) + D(q \parallel p)}{2}. \quad (\text{C.2})$$

Comparing the lower bound (A.5) on Le Cam’s distance with the upper bound (27), we conclude that:

$$e^{d_{\text{JS}}^2(p, q)} - 1 \leq \tanh \frac{D(p, q)}{2}. \quad (\text{C.3})$$

Using  $\tanh x = 2/(1 + e^{-2x}) - 1$ , this can be rewritten as

$$d_{\text{JS}}^2(p, q) \leq \ln \frac{2}{1 + e^{-D(p, q)}} \quad (\text{C.4})$$

which is the main result of [54].

### Appendix D. Switching systems and bang–bang controls

When controlling the parameters of a dynamics in order to elicit a certain behavior, it is common to adopt a strategy of abrupt switching (at well-chosen time instants) between two or several continuous-time dynamics. For instance we can consider a time-varying Markov chain which switches between two transition rate matrices. One reason for such a switching control strategy is the simplicity of implementation, where one has to simply switch on or off a voltage source, a magnetic flux, a reactant influx, a heater, etc.

A deeper reason is to be found in optimal control theory. Even if we can modulate freely the external force with a real number  $u(t)$ , some various fundamental or practical reasons often constraint the values of  $u$  to an interval  $[u_{\min}, u_{\max}]$ . Assume we want to optimize an objective for the trajectory of the system, for instance the time needed to reach a desired state (or set of states), over all possible input signals  $t \mapsto u(t) \in [u_{\min}, u_{\max}]$ . Finding this optimal control signal  $u(t)$  results in a variational problem whose solution is, in many cases, a so-called *bang–bang control* (or on-off control) switching abruptly between the extremal values  $u_{\min}$  and  $u_{\max}$  at well-chosen time instants. Typically, minimum-time control problems are well-known to often result in bang–bang controls (we refer to textbooks such as [55] for precise mathematical statements, e.g. the Pontryagin maximum principle). A didactical example is the following: suppose that we want to drive a car from point A (at zero speed) to point B (at zero speed) along a line: how to proceed in minimum time? The answer is accelerate as much as is possible for the car, then at some carefully calculated moment, brake as much as possible so as to arrive at B at zero speed. A bit switch is another example: if the input voltage is constrained

to be in the interval  $[u_{\min}, u_{\max}]$ , then the abrupt bang–bang switching is, generically speaking, the fastest way to write a bit into the memory. Thus the results illustrated on the bit switch have a more general value for systems operated ‘as fast as possible’, hence in an abrupt switching way.

## Appendix E. Distinguishing the arrow of time: low error requires high entropy production

In [22], the authors consider the following decision problem: we are given a trajectory of a stochastic system, which is presented either in the forward time direction or in the backward time direction (with probability 1/2), and we must make a guess on which direction is being presented to us. This problem offers another application of the main formula (36), in which  $\Omega$  is regarded as a set of trajectories of a stochastic system. We request that this set is closed under time-reversal. The trajectories can be on fixed time-interval, or of variable length (determined by a random starting and stopping time). The two distributions  $p$  and  $q$  of interest are the forward path probability and the reverse path probability, respectively. Now  $\Delta\sigma$  is, under some circumstances (for instance a constant, or time-symmetric protocol), the entropy production along the path, or a lower bound on it. The probability of error in the maximum likelihood decision is  $\epsilon$ . Our formula (36) once again shows that a small probability of error requires a large entropy production. While formally identical to Equation (9) in [22], (36) holds under very general circumstances, e.g. independently of the choice of a certain stopping criterion and for non-Markovian dynamics.

## Appendix F. More classical SLs

### F.1. Hölder’s inequality

Instead of using the bound (66), one could use as well Cauchy–Schwartz inequality to separate the thermodynamic from the kinetic factors appearing in (65). Hölder’s inequality can also be used,

$$\left| \int fg dt \right| \leq \left( \int |f|^k dt \right)^{1/k} \left( \int |g|^\ell dt \right)^{1/\ell}, \quad (\text{F.1})$$

for any  $k, \ell \in [1, +\infty]$  such that  $1/k + 1/\ell = 1$  and functions  $f, g$  so that the integrals make sense. In fact, this inequality generalizes both Cauchy–Schwartz inequality ( $k = \ell = 2$ ) and (66) ( $k = \infty, \ell = 1$ ). In this way, every choice of  $k, \ell$  leads to a variant classical SL of (67). The same remark holds for (75).

### F.2. More distances

Many distances can be defined on the space of probability distributions, all with their respective applications and interpretations. In addition to Le Cam’s distance  $d_{\text{LeCam}}$  and Jensen–Shannon’s distance  $d_{\text{JS}}$ , we can also define Hellinger’s distance,

$$d_{\text{Hell}}(p, q) = \frac{1}{2} \|\sqrt{p} - \sqrt{q}\|_2, \quad (\text{F.2})$$

or Fisher’s distance  $d_{\text{Fisher}}$ , which is the length in Fisher’s information metric sense of the shortest path between two distributions. It turns out that these four distances essentially coincide

locally. Indeed, if  $q = p + \Delta p$  with  $\Delta p \ll 1$ , then

$$d_{\text{LeCam}}^2(p, q) \approx 2d_{\text{JS}}^2(p, q) \approx 2d_{\text{Hell}}^2(p, q) \approx \frac{1}{2}d_{\text{Fisher}}^2(p, q) \approx \frac{1}{4}\sum_{\omega} \frac{\Delta p^2}{p}. \quad (\text{F.3})$$

Thus one can replace Le Cam's distance  $d_{\text{LeCam}}$  with  $\sqrt{2}d_{\text{Hell}}$ , or with  $\sqrt{2}d_{\text{JS}}$ , or with  $\sqrt{1/2}d_{\text{Fisher}}$  in SLs (65), (67) and (68).

## ORCID iDs

Gianmaria Falasco  <https://orcid.org/0000-0003-2870-2787>

## References

- [1] Uffink J and Van Lith J 1999 *Found. Phys.* **29** 655–92
- [2] Puglisi A, Sarracino A and Vulpiani A 2017 *Phys. Rep.* **709** 1–60
- [3] Barato A C and Seifert U 2015 *Phys. Rev. Lett.* **114** 158101
- [4] Gingrich T R, Horowitz J M, Perunov N and England J L 2016 *Phys. Rev. Lett.* **116** 120601
- [5] Polettini M, Lazarescu A and Esposito M 2016 *Phys. Rev. E* **94** 052104
- [6] Horowitz J M and Gingrich T R 2017 *Phys. Rev. E* **96** 020103
- [7] Proesmans K and Van den Broeck C 2017 *Europhys. Lett.* **119** 20001
- [8] Gingrich T R and Horowitz J M 2017 *Phys. Rev. Lett.* **119** 170601
- [9] Dechant A and ichi Sasa S 2018 *J. Stat. Mech.* **063209**
- [10] Potts P P and Samuelsson P 2019 *Phys. Rev. E* **100** 052137
- [11] Hasegawa Y and Van Vu T 2019 *Phys. Rev. Lett.* **123** 110602
- [12] Van Vu T and Hasegawa Y 2020 *Phys. Rev. Res.* **2** 013060
- [13] Liu K, Gong Z and Ueda M 2020 *Phys. Rev. Lett.* **125** 140602
- [14] Falasco G, Esposito M and Delvenne J C 2020 *New J. Phys.* **22** 053046
- [15] Horowitz J M and Gingrich T R 2019 *Nat. Phys.* **16** 15–20
- [16] Dechant A and Sasa S i 2020 *Proc. Natl Acad. Sci. USA* **117** 6430–6
- [17] Van den Broeck C and Esposito M 2015 *Physica A* **418** 6–16
- [18] Shiraishi N, Funo K and Saito K 2018 *Phys. Rev. Lett.* **121** 070601
- [19] Vo V T, Van Vu T and Hasegawa Y 2020 *Phys. Rev. E* **102** 062132
- [20] Owen J A, Gingrich T R and Horowitz J M 2020 *Phys. Rev. X* **10** 011066
- [21] Proesmans K and Horowitz J M 2019 *J. Stat. Mech.* **054005**
- [22] Roldán E, Neri I, Dörpinghaus M, Meyr H and Jülicher F 2015 *Phys. Rev. Lett.* **115** 250602
- [23] Lehmann E L and Casella G 2006 *Theory of Point Estimation* (Berlin: Springer)
- [24] Hatano T and Sasa S 2001 *Phys. Rev. Lett.* **86** 3463–6
- [25] Esposito M and Van den Broeck C 2010 *Phys. Rev. E* **82** 011143
- [26] Esposito M and Van den Broeck C 2011 *Europhys. Lett.* **95** 40004
- [27] Sivak D A and Crooks G E 2012 *Phys. Rev. Lett.* **108** 150601
- [28] Parrondo J M, Horowitz J M and Sagawa T 2015 *Nat. Phys.* **11** 131–9
- [29] Hasegawa Y and Van Vu T 2019 *Phys. Rev. E* **99** 062126
- [30] Di Terlizzi I and Baiesi M 2018 *J. Phys. A: Math. Theor.* **52** 02LT03
- [31] Le Cam L 2012 *Asymptotic Methods in Statistical Decision Theory* (Berlin: Springer)
- [32] Vincze I 1981 On the concept and measure of information contained in an observation *Contributions to Probability* (New York: Academic Press) pp 207–14
- [33] Gilardoni G L 2006 *C. R. Math.* **343** 763–6
- [34] Bérut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R and Lutz E 2012 *Nature* **483** 187–9
- [35] Jun Y, Gavrilov M and Bechhoefer J 2014 *Phys. Rev. Lett.* **113** 190601
- [36] Freitas N, Delvenne J C and Esposito M 2021 Stochastic thermodynamics of nonlinear electronic circuits: a realistic framework for computing around  $kT$  *Phys. Rev. X* **11** 031064

- [37] Freitas N, Proesmans K and Esposito M 2021 arXiv:2103.01184
- [38] Kish L 2004 *IEE Proc., Circuits Devices Syst.* **151** 190–4
- [39] Shanbhag N 2002 Reliable and energy-efficient digital signal processing *Proc. 39th Annual Design Automation Conf.* pp 830–5
- [40] Alaghi A, Qian W and Hayes J P 2017 *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **37** 1515–31
- [41] Duembgen L 2010 arXiv:1012.2063
- [42] Proesmans K, Ehrich J and Bechhoefer J 2020 *Phys. Rev. Lett.* **125** 100602
- [43] Baiesi M and Falasco G 2015 *Phys. Rev. E* **92** 042162
- [44] Bretagnolle J and Huber C 1979 *Z. Wahrscheinlichkeit.* **47** 119–37
- [45] Falasco G and Baiesi M 2016 *New J. Phys.* **18** 043039
- [46] Falasco G and Baiesi M 2016 *Europhys. Lett.* **113** 20005
- [47] Falasco G, Baiesi M, Molinaro L, Conti L and Baldovin F 2015 *Phys. Rev. E* **92** 022129
- [48] Falasco G and Esposito M 2020 *Phys. Rev. Lett.* **125** 120604
- [49] Carlson J M and Doyle J 2002 *Proc. Natl Acad. Sci.* **99** 2538–45
- [50] Ames A D and Doyle J C 2008 Complexity and fragility in stability for linear systems 2008 *American Control Conf.* (IEEE) pp 1630–7
- [51] Falasco G, Cossetto T, Penocchio E and Esposito M 2019 *New J. Phys.* **21** 073005
- [52] Forastiere D, Falasco G and Esposito M 2020 *J. Chem. Phys.* **152** 134101
- [53] Polettini M, Falasco G and Esposito M 2021 arXiv:2106.00425
- [54] Crooks G 2008 Inequalities between the Jenson–Shannon and Jeffreys divergences *Tech. Rep. Note 004* Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
- [55] Kirk D E 2004 *Optimal Control Theory: An Introduction* (New York: Academic Press)