

Entropy production as correlation between system and reservoir

Massimiliano Esposito¹, Katja Lindenberg²
and Christian Van den Broeck³

¹ Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Code Postal 231, Campus Plaine, B-1050 Brussels, Belgium

² Department of Chemistry and Biochemistry and BioCircuits Institute, University of California, San Diego, La Jolla, CA 92093-0340, USA

³ Department of WNI, Hasselt University, B 3590 Diepenbeek, Belgium
E-mail: mesposit@ulb.ac.be

New Journal of Physics **12** (2010) 013013 (10pp)

Received 1 October 2009

Published 15 January 2010

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/12/1/013013

Abstract. We derive an exact (classical and quantum) expression for the entropy production of a finite system placed in contact with one or several finite reservoirs, each of which is initially described by a canonical equilibrium distribution. Although the total entropy of system plus reservoirs is conserved, we show that system entropy production is always positive and is a direct measure of system–reservoir correlations and/or entanglements. Using an exactly solvable quantum model, we illustrate our novel interpretation of the Second Law in a microscopically reversible finite-size setting, with strong coupling between the system and the reservoirs. With this model, we also explicitly show the approach of our exact formulation to the standard description of irreversibility in the limit of a large reservoir.

Contents

1. Introduction	2
2. An exact form for entropy production	3
3. Discussion	4
3.1. Meaning of entropy production	5
3.2. Recurrences	6
3.3. Connection to the First Law	6
3.4. Entropy production in the weak coupling limit	6
4. Model system	7
5. Conclusions	8
Acknowledgments	9
References	9

1. Introduction

Starting with the ground-breaking work of Boltzmann, there have been numerous attempts to construct a microscopic derivation of the Second Law. The main difficulty is that the prime microscopic candidate for entropy, namely the von Neumann entropy $S = -\text{Tr} \rho \ln \rho$, with ρ the density matrix of the total or compound system, is a constant in time by virtue of Liouville's theorem. Related difficulties are the time reversibility of the microscopic laws and recurrences of the micro-states. A common way to bypass these difficulties is to introduce irreversibility in an *ad hoc* way, for example by reasoning that the system is in contact with idealized infinitely large heat reservoirs. Nevertheless, as was realized early on by Onsager, a consistent description of the resulting irreversible behavior still carries the undiluted imprint of the underlying time reversibility and Liouville's theorem for the system. Examples are the symmetry of the Onsager coefficients and the fluctuation dissipation theorem. As examples of more recent discussions, we cite the results on work theorems and fluctuation theorems [1]–[3]. Even more relevant to the question pursued here, we cite the microscopic expression for entropy production as the breaking, in a statistical sense, of the arrow of time [4]–[10]. We also mention that significant effort has been devoted to a detailed description and understanding of the interaction with heat reservoirs, in particular the difficulties of dealing with the case of strong coupling [11, 12].

In this paper, we show that the problem of entropy production can be addressed within a microscopically exact description of a finite system, without resorting to infinitely large heat reservoirs and without any assumption of weak coupling. Whereas the von Neumann entropy of system plus reservoirs is conserved, the entropy production of the system is always positive, even though it displays oscillations and recurrences typical of the finite total system. Interestingly, this entropy production is expressed in terms of the correlations and/or entanglements between system and reservoirs, so that its positivity can be explained by a corresponding negative entropy contribution contained in the correlations and/or entanglements with the reservoirs. As the size of the reservoirs increases, the recurrences die out, the negative entropy contribution is diluted in an intricate way over the increasing number of correlations with reservoir degrees of freedom, and the entropy production of the system itself approaches the standard thermodynamic form. We will illustrate this novel interpretation of the Second Law on an exactly solvable model, namely, a spin interacting with an N -level quantum system via

a random matrix coupling. We focus on the derivation for the quantum case, but the analogous treatment for the classical system is straightforward.

2. An exact form for entropy production

The setup is as follows. We consider one or several finite quantum systems r which play the role of finite-size heat reservoirs. Accordingly, their density matrices $\rho_r(t)$ at initial time $t = 0$ are assumed to be of canonical equilibrium form,

$$\rho_r(0) = \rho_r^{\text{eq}} = \exp(-\beta_r H_r) / Z_r. \quad (1)$$

Here β_r , H_r and Z_r are the corresponding inverse temperature at $t = 0$ (Boltzmann's constant k_B is set equal to 1), the Hamiltonian, and the partition function at $t = 0$. Being reservoir systems, it is natural to assume that their Hamiltonians H_r are time independent. At time $t = 0$, we connect a finite quantum system s , characterized by Hamiltonian $H_s(t)$ and density matrix $\rho_s(t)$, to the reservoirs by switching on an interaction Hamiltonian $V(t)$. The initial state of the compound system, characterized by the density matrix $\rho(t)$, does not display any entanglement or correlation,

$$\rho(0) = \rho_s(0) \prod_r \rho_r^{\text{eq}}. \quad (2)$$

Correlations and/or entanglements do develop in the subsequent time evolution of $\rho(t)$, which obeys Liouville's equation for the total Hamiltonian

$$H(t) = H_s(t) + \sum_r H_r + V(t). \quad (3)$$

Note that in addition to the issue of relaxation of a system in contact with a reservoir, this scenario includes ingredients for the study of a driven system via time dependence of the system's Hamiltonian, as well as that of a non-equilibrium steady state, which can be realized in view of the presence of several heat reservoirs. In fact, the above construct can easily be generalized to include particle reservoirs described via grand-canonical distributions. This would allow the consideration of particle flows in addition to heat flows.

We are primarily interested in the occurrence and characterization of irreversible behavior in the system. We thus focus our attention on entropy $S(t)$ of the system,

$$S(t) \equiv -\text{Tr}_s \rho_s(t) \ln \rho_s(t), \quad (4)$$

where $\rho_s(t)$ is the trace of $\rho(t)$ over the degrees of freedom of all the reservoirs. Contrary to the total von Neumann entropy, the entropy of the system is in general a function of time, because technically speaking the dynamics of $\rho_s(t)$ is not unitary. More to the point for the ensuing discussion, we note that from the thermodynamic point of view we are dealing with an energetically open system. We now show that it is precisely the time invariance of the total von Neumann entropy that induces a natural separation of the entropy change of the system into separate contributions from an entropy flow and an entropy production. Using

$$-\text{Tr} \rho(t) \ln \rho(t) = -\text{Tr} \rho(0) \ln \rho(0) = -\text{Tr}_s \rho_s(0) \ln \rho_s(0) - \sum_r \text{Tr}_r \rho_r^{\text{eq}} \ln \rho_r^{\text{eq}}, \quad (5)$$

we find for the entropy change of the system

$$\begin{aligned} \Delta S(t) &= S(t) - S(0) = -\text{Tr}\rho(t) \ln \rho_s(t) + \text{Tr}\rho(t) \ln \rho(t) - \sum_r \text{Tr}_r \rho_r^{\text{eq}} \ln \rho_r^{\text{eq}} \\ &= -\text{Tr}\rho(t) \ln \left\{ \rho_s(t) \prod_r \rho_r^{\text{eq}} \right\} + \text{Tr}\rho(t) \ln \rho(t) + \sum_r \text{Tr}_r [\rho_r(t) - \rho_r^{\text{eq}}] \ln \rho_r^{\text{eq}}. \end{aligned} \quad (6)$$

We conclude that the change in the entropy of the system can be written in standard thermodynamic form [13]–[16]

$$\Delta S(t) = \Delta_i S(t) + \Delta_e S(t). \quad (7)$$

Entropy flow, representing the reversible contribution to the system entropy change due to heat exchanges, is identified as the last term in (6). After some manipulation using the explicit form of ρ_r^{eq} , it can be written in standard thermodynamic form

$$\Delta_e S(t) = \sum_r \beta_r Q_r(t), \quad (8)$$

where the heat flowing from reservoir r is

$$Q_r(t) \equiv \langle H_r \rangle_0 - \langle H_r \rangle_t. \quad (9)$$

We use the notation $\langle \bullet \rangle_t \equiv \text{Tr}[\rho(t)\bullet]$. Of particular interest is the resulting expression for the entropy production,

$$\Delta_i S(t) \equiv D \left[\rho(t) \parallel \rho_s(t) \prod_r \rho_r^{\text{eq}} \right], \quad (10)$$

which represents the irreversible contribution to the entropy change of the system. Here, $D[\rho \parallel \rho']$ is the quantum relative entropy between two density matrices ρ and ρ' ,

$$D[\rho \parallel \rho'] \equiv \text{Tr}\rho \ln \rho - \text{Tr}\rho \ln \rho'. \quad (11)$$

It has the following important properties [17, 18]. The relative entropy is positive, and equal to zero only when the two matrices are identical. We thus conclude that the entropy production introduced above is indeed a positive quantity, $\Delta_i S(t) \geq 0$, and vanishes only when the system and reservoirs are totally decorrelated. Furthermore, relative entropy is a measure of the ‘distance’ between two density matrices. Hence, as mentioned earlier, entropy production explicitly expresses how ‘far’ the actual state $\rho(t)$ of the total system is from the decorrelated/disentangled product state $\rho_s(t) \prod_r \rho_r^{\text{eq}}$.

We have thus derived in a Hamiltonian framework the central thermodynamic relation, (7) with (8) and (10), for the entropy evolution of a system out of equilibrium.

3. Discussion

To further clarify the significance of our central result (7), we make a number of additional comments.

3.1. Meaning of entropy production

Firstly, we note that the sum of the system and reservoir entropies is not the entropy of the total system because it lacks the entropy contribution contained in the correlations and entanglement between the system and the reservoirs. We thus define the correlation entropy, $S_c(t)$, by

$$S(t) + \sum_r S_r(t) + S_c(t) = -\text{Tr} \rho(t) \ln \rho(t), \quad (12)$$

where the entropy of the reservoir r is

$$S_r(t) \equiv -\text{Tr}_r \rho_r(t) \ln \rho_r(t). \quad (13)$$

We note that $S_c(0) = 0$ because no initial correlations are present (2). Because the entropy of the total system is conserved, we find that

$$S_c(t) = -\Delta S(t) - \sum_r \Delta S_r(t). \quad (14)$$

Using definition (11), we verify that

$$D \left[\rho(t) \parallel \rho_s(t) \prod_r \rho_r(t) \right] = \Delta S(t) + \sum_r \Delta S_r(t) \geq 0, \quad (15)$$

so that from (14) and (15) we conclude that

$$S_c(t) = -D \left[\rho(t) \parallel \rho_s(t) \prod_r \rho_r(t) \right] \leq 0. \quad (16)$$

The entropy contribution contained in the correlation is thus always negative or zero. Using (10), we see that in the thermodynamic limit, when the reservoirs can be assumed to remain at equilibrium, i.e. $\rho_r(t) = \rho_r^{\text{eq}}$, the correlation entropy becomes equal to minus the entropy production $S_c(t) = -\Delta_i S(t)$. Furthermore, since

$$\Delta_i S(t) + S_c(t) = -\sum_r \beta_r Q_r(t) - \sum_r \Delta S_r(t) = \sum_r D[\rho_r(t) \parallel \rho_r^{\text{eq}}] \geq 0, \quad (17)$$

we find that the entropy production is an upper bound to minus the always negative correlation entropy, i.e.

$$\Delta_i S(t) \geq -S_c(t) = \Delta S(t) + \sum_r \Delta S_r(t) \geq 0. \quad (18)$$

Neglecting a part of the negative entropy contribution contained in the correlation, $\Delta_i S(t) + S_c(t)$, is actually quite natural from an operational point of view: while one has full microscopic access to the system's properties, one only controls or measures the energy and no other properties of the reservoir. In this sense, the above procedure leading to the central thermodynamic relation, (7) with (8) and (10), can be viewed as a coarse graining operation that retains the full microscopic description of the system but reduces the reservoirs plus correlations to an idealized heat reservoir description. In the limit of large reservoirs, it is likely that this latter description deviates very little from the actual one. The reservoir entropy can then be approximated by minus the entropy flow $\sum_r \Delta S_r(t) \approx -\Delta_e S(t) = -\sum_r \beta_r Q_r(t)$. Concerning the correlations, one expects that they will be diluted over the exponentially many higher order correlations, becoming in effect irretrievable so that $S_c(t) \approx -\Delta_i S(t)$. Furthermore, this will happen exponentially fast in time if the reservoirs display non-integrable, chaotic properties.

3.2. Recurrences

Secondly, while $\Delta_i S(t)$ is a positive quantity, it does not increase monotonically in time. In fact, oscillations are bound to arise in view of recurrences in the state of the finite total system. In this respect, it is important to stress that we consider the entropy production starting from the natural but specific initial condition (2). The transient decreases of $\Delta_i S(t)$ can be interpreted as the reappearance of the negative entropy, hidden in the correlations, as system and reservoir transiently return to states close to this decoupled initial state. In the limit of large reservoirs, recurrences will become less and less likely, and $\Delta_i S(t)$ is expected to converge to a convex monotonically increasing function of t .

3.3. Connection to the First Law

Thirdly, we discuss the appropriate definition of work and free energy in a driven system strongly coupled to a heat reservoir. We consider the case of a single reservoir at temperature $T = \beta^{-1}$, for convenience dropping the subscript r . Using the fact that $\text{Tr}H(t)\dot{\rho}(t) = 0$, the work done on the total system can be written as

$$W \equiv \langle H(t) \rangle_t - \langle H(0) \rangle_0 = \int_0^t d\tau \text{Tr}(\dot{H}_s(\tau) + \dot{V}(\tau))\rho(\tau). \quad (19)$$

Using $\text{Tr}(H_s(t) + V(t))\dot{\rho}(t) = -\text{Tr}H_r\dot{\rho}(t)$, we can rewrite heat (9) as

$$Q(t) = \int_0^t d\tau \text{Tr}(H_s(\tau) + V(\tau))\dot{\rho}(\tau). \quad (20)$$

The change of the energy of the system, including the contribution of the interaction term, reads

$$\Delta U(t) \equiv \langle (H_s(t) + V(t)) \rangle_t - \langle (H_s(0) + V(0)) \rangle_0 \quad (21)$$

and can be written, in accordance with the First Law, as the sum of work and heat,

$$\Delta U(t) = W(t) + Q(t). \quad (22)$$

Next, introducing the non-equilibrium free energy

$$\Delta F(t) \equiv \Delta U(t) - T \Delta S(t), \quad (23)$$

we can rewrite expression (7) for entropy production in the standard thermodynamic form for a driven system in contact with a heat reservoir,

$$T \Delta_i S(t) = W(t) - \Delta F(t) \geq 0. \quad (24)$$

This expression is exact. If we assume that the total system relaxes to a final canonical equilibrium at temperature β^{-1} , this non-equilibrium free energy difference reduces to the equilibrium expression identified in the context of the work theorem in both the weak coupling [2, 4, 19] and strong coupling regimes [11, 12].

3.4. Entropy production in the weak coupling limit

Finally, we discuss the connection with the following alternative definition for entropy production, proposed in open quantum system theory [18]:

$$\Delta_i \bar{S}(t) \equiv D[\rho_s(0) || \rho_s^{\text{eq}}] - D[\rho_s(t) || \rho_s^{\text{eq}}] = \Delta S(t) - \Delta_e \bar{S}(t). \quad (25)$$

Entropy flow is now defined as

$$\Delta_e \bar{S}(t) \equiv \beta(\langle H_s \rangle_t - \langle H_s \rangle_0) \quad (26)$$

and $\rho_s^{\text{eq}} = \exp(-\beta H_s)/Z_s$. To compare this expression with our definition (10) for entropy production, we note that total energy is conserved by the dynamics, $\langle H \rangle_t = \langle H \rangle_0$, and hence

$$\Delta_i S(t) = \Delta_i \bar{S}(t) - \beta(\langle V \rangle_t - \langle V \rangle_0) \geq 0. \quad (27)$$

The two definitions thus differ by the interaction term, which vanishes in the limit of either weak coupling or high temperature. Definition (25) has the obvious advantage of being exclusively expressed in terms of the system density matrix, whereas our definition (10) requires the total density matrix. However, we will show that, contrary to our expression, (25) is not always a positive quantity. The positivity of (25) can be proven when ρ_s^{eq} is the stationary solution of the reduced system dynamics [18]. This will generically be the case in the weak-interaction large-reservoir limit, i.e. precisely when the interaction term in (27) can be neglected and (25) becomes identical to (10). An even stronger statement can be made when, in the same limit, the system dynamics can be described by a Markovian quantum master equation of the form $\dot{\rho}_s(t) = \mathcal{L}\rho_s(t)$, where \mathcal{L} is a superoperator satisfying $\mathcal{L}\rho_s^{\text{eq}} = 0$. An example of such \mathcal{L} is the Redfield superoperator that can be derived using second-order perturbation theory in the system–reservoir interaction [18, 20]. Under these conditions, it is known that entropy production is a convex functional of the system density matrix [18], with a positive rate of entropy production

$$\frac{d}{dt} \Delta_i S(t) \approx \frac{d}{dt} \Delta_i \bar{S}(t) \geq 0. \quad (28)$$

4. Model system

We will now illustrate the above findings in a two-level quantum spin coupled to an N -level reservoir via a random matrix. The total Hamiltonian reads

$$H = \frac{\Delta}{2} \sigma_z + H_r + \lambda \sigma_x R. \quad (29)$$

$\sigma_{x,z}$ are the well-known Pauli matrices. The reservoir Hamiltonian H_r is a diagonal matrix with N equally spaced eigenvalues between -0.5 and 0.5 . The coupling matrix is $R = X/\sqrt{8N}$, where X is a Gaussian orthogonal random matrix of size N with probability density proportional to $\exp(-\frac{1}{4}\text{Tr}X^2)$ [21, 22]. This model is similar to the spin-Gaussian orthogonal random matrices (GORM) model of Esposito and Gaspard [23]. The system is initially assumed to be in the pure lower energy state $\rho_s(0) = |0\rangle\langle 0|$, where $\sigma_z|0\rangle = -|0\rangle$, and the reservoir is initially in a canonical equilibrium state at temperature β^{-1} . In the weak-coupling large-reservoir limit, the resulting Redfield equation leads to the following closed relaxation equation for the z -component of the spin ($\hbar = 1$):

$$\langle \sigma_z \rangle_t = \langle \sigma_z \rangle_{\text{eq}} + (\langle \sigma_z \rangle_0 - \langle \sigma_z \rangle_{\text{eq}}) e^{-\gamma t}, \quad (30)$$

where

$$\gamma = 2\pi\lambda^2(\tilde{\alpha}(\Delta) + \tilde{\alpha}(-\Delta)), \quad \langle \sigma_z \rangle_{\text{eq}} = \frac{\tilde{\alpha}(-\Delta) - \tilde{\alpha}(\Delta)}{\tilde{\alpha}(-\Delta) + \tilde{\alpha}(\Delta)}. \quad (31)$$

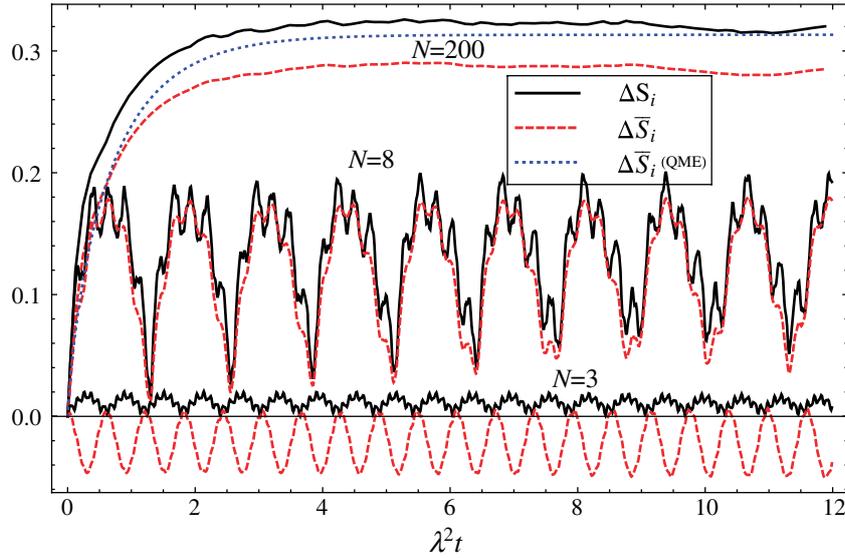


Figure 1. ΔS_i (resp. $\Delta \bar{S}_i$) is entropy production (10) [(25)] calculated using the exact numerical dynamics. $\Delta \bar{S}_i$ (QME) is the irreversible entropy production (25) calculated using the Redfield equation (30). Parameters are $\Delta = 0.1$, $\beta = 10$ and $\lambda = 0.1$.

Here $\tilde{\alpha}(\omega)$ is the Fourier transform of the reservoir correlation function

$$\alpha(t) \equiv \text{Tr}_r \rho_r^{\text{eq}} \exp[iH_r t] R \exp[-iH_r t] R, \quad (32)$$

$$\tilde{\alpha}(|\omega|) = \frac{1}{16} \frac{e^{-\beta/2} e^{\beta|\omega|} - e^{\beta/2}}{e^{-\beta/2} + e^{\beta/2}} = e^{\beta|\omega|} \tilde{\alpha}(-|\omega|). \quad (33)$$

The x - and y -components of the spin evolve independently of the z -component, and are zero for our initial condition.

We are now in a position to compare our definition (10) of entropy production with definition (25) calculated from the exact dynamics as well as from Redfield theory. This is accomplished through an exact numerical solution of our model for finite N . The results are summarized in figure 1. Note that we show single realizations of the random matrix. For small N , we observe a pronounced oscillatory behavior of our entropy production (10) and even near-recurrences very close to zero. Whereas our entropy production always remains positive, entropy production (25) can be negative for small values of N , which is clearly not acceptable. In the limit of a large reservoir ($N \rightarrow \infty$), both expressions converge to one another and coincide with the positive and convex irreversible entropy production predicted by the Redfield equation.

5. Conclusions

We conclude that (10) is a proper definition for entropy production, one that remains valid for a small system strongly coupled to small reservoirs. In the limit of large reservoirs, it converges to a convex irreversible entropy, coinciding with the familiar definition of entropy

production [18] for the quantum master equation. Our identification of entropy production (10) within an exact microscopic framework vindicates the description of irreversibility as a property of open systems, with the entropy production rather than entropy of the total system playing the central role. The microscopic origin of entropy production, explained in terms of correlations established between the system and its reservoirs, is reminiscent of Boltzmann's Stosszahlansatz. However, our analysis of the micro-dynamics and the identification of the entropy production of the system are exact. The appearance of irreversibility as the omission, in the reservoirs, of its correlations with the system provides a natural, precise and transparent interpretation of the Second Law.

Acknowledgments

ME is supported by FNRS Belgium (chargé de recherches) and by the Government of Luxembourg (Bourse de formation recherches). KL and C VdB gratefully acknowledge the support of the US National Science Foundation through grant no. PHY-0855471.

References

- [1] Gallavotti G and Cohen E G D 1995 Dynamical ensembles in nonequilibrium statistical mechanics *Phys. Rev. Lett.* **74** 2694
- [2] Jarzynski C 1997 Nonequilibrium equality for free energy differences *Phys. Rev. Lett.* **78** 2690
- [3] Seifert U 2005 Entropy production along a stochastic trajectory and an integral fluctuation theorem *Phys. Rev. Lett.* **95** 040602
- [4] Kawai R, Parrondo J M R and Van den Broeck C 2007 Dissipation: the phase-space perspective *Phys. Rev. Lett.* **98** 080602
- [5] Gomez-Marin A, Parrondo J M R and Van den Broeck C 2008 The footprints of irreversibility *Europhys. Lett.* **82** 50002
- [6] Gaspard P 2004 Time-reversed dynamical entropy and irreversibility in Markovian random processes *J. Stat. Phys.* **117** 599
- [7] Andrieux D, Gaspard P, Ciliberto S, Garnier N, Joubaud S and Petrosyan A 2007 Entropy production and time asymmetry in nonequilibrium fluctuations *Phys. Rev. Lett.* **98** 150601
- [8] Esposito M, Harbola U and Mukamel S 2009 Nonequilibrium fluctuations, fluctuation theorems and counting statistics in quantum systems *Rev. Mod. Phys.* **81** 1665
- [9] Parrondo J M R, Van den Broeck C and Kawai R 2009 Entropy production and the arrow of time *New J. Phys.* **11** 073008
- [10] Jarzynski C 2006 Rare events and the convergence of exponentially averaged work values *Phys. Rev. E* **73** 046105
- [11] Jarzynski C 2004 Nonequilibrium work theorem for a system strongly coupled to a thermal environment *J. Stat. Mech.* **P09005**
- [12] Campisi M, Talkner P and Hanggi P 2009 Fluctuation theorem for arbitrary open quantum systems *Phys. Rev. Lett.* **102** 210401
- [13] Kondepudi D and Prigogine I 1998 *Modern Thermodynamics* (New York: Wiley)
- [14] de Groot S R and Mazur P 1984 *Non-Equilibrium Thermodynamics* (New York: Dover)
- [15] Prigogine I and Nicolis G 1977 *Self-Organization in Non-Equilibrium Systems* (New York: Wiley)
- [16] Gaspard P 2006 Hamiltonian dynamics, nanosystems and nonequilibrium statistical mechanics *Physica A* **369** 201
- [17] Nielsen M A and Chuang I 2000 *Quantum computation and quantum information* (Cambridge: Cambridge University Press)

- [18] Breuer H-P and Petruccione F 2002 *The Theory of Open Quantum Systems* (Oxford: Oxford University Press)
- [19] Talkner P, Campisi M and Hänggi P 2009 Fluctuation theorems in driven open quantum systems *J. Stat. Mech.* [P02025](#)
- [20] Gaspard P and Nagaoka M 1999 Slippage of initial conditions for the redfield master equation *J. Chem. Phys.* [111 5668](#)
- [21] Mehta M L 2004 *Random Matrices* 3rd edn (New York: Academic)
- [22] Haake F 2001 *Quantum Signatures of Chaos* (Berlin: Springer)
- [23] Esposito M and Gaspard P 2003 Spin relaxation in a complex environment *Phys. Rev. E* [68 066113](#)