Transfer Entropy and Transient Limits of Computation

Mikhail Prokopenko[1,2,3,⋆] and Joseph T. Lizier[1]

[1] *CSIRO Computational Informatics, PO Box 76, Epping, NSW 1710, Australia*
[2] *Department of Computing, Macquarie University, E6A Level 3, Eastern Rd, Macquarie Park, NSW 2113, Australia*
[3] *School of Physics, University of Sydney, Physics Rd, Camperdown NSW 2050, Australia*
⋆*corresponding author: mikhail.prokopenko@csiro.au*

Transfer entropy is a recently introduced information-theoretic measure quantifying directed statistical coherence between spatiotemporal processes, and is widely used in diverse fields ranging from finance to neuroscience. However, its relationships to fundamental limits of computation, such as Landauer's limit, remain unknown. Here we show that in order to increase transfer entropy (predictability) by one bit, heat flow must match or exceed Landauer's limit. Importantly, we generalise Landauer's limit to bidirectional information dynamics for non-equilibrium processes, revealing that the limit applies to prediction, in addition to retrodiction (information erasure). Furthermore, the results are related to negentropy, and to Bremermann's limit and the Bekenstein bound, producing, perhaps surprisingly, *lower* bounds on the computational deceleration and information loss incurred during an increase in predictability about the process. The identified relationships set new computational limits in terms of fundamental physical quantities, and establish transfer entropy as a central measure connecting information theory, thermodynamics and theory of computation.

Transfer entropy [1] was designed to determine the direction of information transfer between two, possibly coupled, processes, by detecting asymmetry in their interactions. It is a Shannon information-theoretic quantity [2–4] which measures a directed relationship between two time-series processes Y and X. Specifically, the transfer entropy $T_{Y \to X}$ measures the average amount of information that states $\mathbf{y_n}$ at time n of the *source* time-series process Y provide about the next states $\mathbf{x_{n+1}}$ of the *destination* time-series process X, in the context of the previous state $\mathbf{x_n}$ of the destination process (see more details in Methods):

$$T_{Y \to X} = \left\langle \log_2 \frac{p(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n})}{p(\mathbf{x_{n+1}} \mid \mathbf{x_n})} \right\rangle. \tag{1}$$

The definition is asymmetric in Y and X, hence the labelling of an information *source* and *destination*. Intuitively, it helps to answer the question *"if I know the state of the source, how much does that help to predict the state transition of the destination?"*.

Following the seminal work of Schreiber [1] numerous applications of transfer entropy have been successfully developed, by capturing information transfer within various domains, such as finance [5], ecology [6], neuroscience [7,8], biochemistry [9], distributed computation [10–12], statistical inference [13], complex systems [14], complex networks [15,16], robotics [17], etc. Interestingly, maxima of transfer entropy were observed to be related to critical behaviour, e.g.,

1

average transfer entropy was observed to maximize on the chaotic side of the critical regime within random Boolean networks [18], and was analytically shown to peak on the disordered side of the phase transition in a ferromagnetic 2D lattice Ising model with Glauber dynamics [19]. Transfer entropy was also found to be high while a system of coupled oscillators was beginning to synchronize, followed by a decline from the global maximum as the system was approaching a synchronized state [16]. Similarly, transfer entropy was observed to be maximized in coherent propagating spatiotemporal structures within cellular automata (i.e., gliders) [10], and self-organizing swarms (cascading waves of motions) [14].

There is growing awareness that information is a physical quantity, with several studies relating various information-theoretic concepts to thermodynamics [20–24], primarily through Landauer's principle [25]. In this paper we report on a physical interpretation of transfer entropy, and its connections to fundamental limits of computation, such as Landauer's limit.

Landauer's principle, dating back to 1961 [25], is a physical principle specifying the lower theoretical limit of energy consumption need for a computation. It associates the logical irreversibility of functions involved in the computation with physical irreversibility, requiring a minimal heat generation per machine cycle for each irreversible function. According to Bennett [20], "any logically irreversible manipulation of information, such as the erasure of a bit . . ., must be accompanied by a corresponding entropy increase in non-information bearing degrees of freedom of the information processing apparatus or its environment". Specifically, the principle states that irreversible destruction of one bit of information results in dissipation of at least $kT \log 2$ J of energy into the environment (i.e. an entropy increases in the environment by this amount — the *Landauer limit*). Here T is the temperature of the computing circuit in kelvins and k is Boltzmann's constant.

We shall consider the non-equilibrium thermodynamics of a physical system X close to equilibrium. At any given moment in time, n, the thermodynamic state $\mathbf{X_n}$ of the physical system is given by a vector $\mathbf{x_n} \in R^d$, comprising d variables, for instance the (local) pressure, temperature, chemical concentrations and so on. A state vector completely describes the physical macrostate as far as predictions of the outcomes of all possible measurements performed on the system are concerned [26]. The state space of the system is the set of all possible states of the system.

The thermodynamic state is generally considered as a fluctuating entity so that conditional probability for a transition from $\mathbf{x_n}$ to $\mathbf{x_{n+1}}$, that is, $p(\mathbf{x_{n+1}}|\mathbf{x_n})$, is a clearly defined property of the system, and can be accurately estimated by a proper sampling procedure. Each macrostate can be realised by a number of different microstates consistent with the given thermodynamic variables. Importantly, in the theory of non-equilibrium thermodynamics close to equilibrium, the microstates belonging to one macrostate $\mathbf{x}$ are equally probable.

A state vector, $\mathbf{y}$, describes a state of some exterior system Y, possibly coupled to the

system represented by X. Due to the presence or lack of such coupling, the time-series processes corresponding to X and Y may or may not be dependent. For a state transition from $\mathbf{x_n}$ to $\mathbf{x_{n+1}}$, we shall say that $\sigma(x)_y$ is the internal entropy production of X *in the context of* some source Y, while $\Delta S(x)_{ext}$ is the entropy production *attributed to* Y, so that (see Methods, (22)):

$$\Delta S(x) = \sigma(x)_y + \Delta S(x)_{ext}, \tag{2}$$

where $\Delta S(x)$ is the total variation of the entropy of system X.

Henceforth, we shall consider two simple examples illustrating entropy dynamics. In both these examples the physical system X is surrounded by the physical system Y (note that X is not a component of Y). The first example is the classical Joule expansion, and the second is compression in the Szilárd engine-like device. These examples are fully described in Methods, illustrating cases with and without external entropy production. A comparison between these two examples shows that, although the resultant entropy change is the same in magnitude $|\Delta S(x)| = k \log 2$ in both cases, the change is brought about differently: (i) for the Joule expansion (of a one molecule gas), $\sigma(x)_y = k \log 2$ and $\Delta S(x)_{ext} = 0$, while (ii) for the Szilárd engine's compression (resetting one bit) $\sigma(x)_y = 0$ and $\Delta S(x)_{ext} = -k \log 2$.

At this stage, we would like to point out two important lessons from this comparison. Firstly, as argued by Bennett [20], "a logically irreversible operation, such as the erasure of a bit or the merging of two paths, may be thermodynamically reversible or not depending on the data to which it is applied". The (computation) paths referred to here are trajectories through the *state-space* or phase portrait of a system: if at time step $n + 1$ we reach a global system state $\mathbf{x}_{n+1}$ with multiple precursor states $\mathbf{x}_n$, then we have irreversibly destroyed information which is non-zero. This information is quantified by the entropy of the previous state conditioned on the next state, i.e., the conditional entropy $h(\mathbf{x_n} \mid \mathbf{x_{n+1}})$ at time step $n + 1$ (see Methods). Bennett elaborates that, if the data being erased is random, its erasure would represent a reversible entropy transfer to the environment, compensating an earlier entropy transfer from the environment during, e.g., a previous isothermal expansion. This, as expected, would make the total work yield of the cycle zero, in obedience to the Second Law. For a bit reset, as for any usual deterministic digital computation, the data is not random, being determined by the device's initial state — this is a crucial difference, as pointed out by Landauer and Bennett.

It is worth pointing out that the laws "of a closed physical system are one-to-one" [22, p.78], meaning that in closed physical systems (or the universe as a whole) computational paths do not merge, and information cannot be destroyed. We can only measure information destruction in *open systems*, and what we measure is a departure of information from this system into the external, unobserved environment, where the destroyed information is offloaded along with energy dissipation. Note that a typical connection between an observed computational system and external environment is the physical representation of the computational system (e.g. bit registers): after all, "information is physical" [27].

Secondly, we note that for the Szilárd engine's compression (resetting the bit), there is a decrease in the thermodynamical entropy of the one molecule gas by $k \log 2$, and so one may argue that there is an increase in predictability about this system transition. It is precisely this intuition that was recently formalized via transfer entropy capturing the external entropy production [28].

Arguably, transfer entropy takes an opposite perspective to information destruction, focussing on ability for prediction in coupled systems, rather than uncertainty in retrodiction. In our recent work [28], transfer entropy has been precisely interpreted thermodynamically. The proposed thermodynamic interpretation of transfer entropy near equilibrium used the specialised Boltzmann's principle, and related conditional probabilities to the probabilities of the corresponding state transitions. This in turn characterised transfer entropy as a difference of two entropy rates: the rate for a resultant transition and another rate for a possibly irreversible transition within the system affected by an additional source. Then it was shown that this difference, the local transfer entropy, is proportional to the external entropy production, possibly due to irreversibility. In the following sections we revisit the main elements of this approach, leading to new fundamental connections with Landauer's principle, Bremermann's limit, the Bekenstein bound, as well as Massieu-Planck thermodynamic potential (free entropy).

Results

Preliminary results: transfer entropy as external entropy production. Supported by the background described in Methods, specifically the assumptions (23) – (24) and the expressions (18) – (20), transfer entropy can be interpreted via transitions between states [28]:

$$t_{Y \to X}(n+1) = \log_2 \frac{Z_1}{Z_2} + \frac{1}{k \log 2} \left(- \left(S(\mathbf{x_{n+1}}) - S(\mathbf{x_n}) \right) + \sigma(x)_y \right). \tag{3}$$

When one considers a small fluctuation near an equilibrium, $Z_1 \approx Z_2$, as the number of microstates in the macrostates does not change much. This removes the additive constant. Using the expression for entropy production (2), one obtains

$$t_{Y \to X}(n+1) = -\frac{\Delta S(x)_{ext}}{k \log 2}. \tag{4}$$

If $Z_1 \neq Z_2$, the relationship includes some additive constant $\log_2 \frac{Z_1}{Z_2}$.

That is, the transfer entropy is proportional to the external entropy production, brought about by the source of irreversibility Y. The opposite sign reflects the different direction of entropy production attributed to the source Y: when $\Delta S(x)_{ext} > 0$, i.e. the entropy due to interactions with the surroundings increased during the transition in X, then the local transfer entropy is negative, and the source misinforms about the macroscopic state transition. When, on the other hand, $\Delta S(x)_{ext} < 0$, i.e. some entropy produced during the transition in X dissipated to the exterior, then the local transfer entropy is positive, and better predictions can be made about the macroscopic state transitions in X if source Y is measured.

4

Turning to our examples (see Methods), we note that (i) for the Joule expansion $\Delta S(x)_{ext} = 0$, and so, according to equation (4), $t_{Y \to X} = 0$ as well, as the transition is adiabatic and X and Y are independent processes, while (ii) for the Szilárd engine's compression $\Delta S(x)_{ext} = -k \log 2$, and so equation (4) yields $t_{Y \to X} = 1$. That is, in the latter case of resetting one bit, the decrease in the thermodynamical entropy of the one molecule gas by $k \log 2$ is accompanied by an increase in predictability about this state transition, by one bit precisely: this increase is captured by the transfer entropy from the exterior heat bath Y to the container X.

It is important to realize that local transfer entropy may increase, indicating an increase of predictability about a transition, not necessarily only when there is an irreversible operation, such as a bit reset. In other words, predictability about a transition may be increased for a wider range of processes. The transfer entropy quantifies then the *extent* by how much the predictability is increased. Furthermore, in general, there is a distinction between *retrodiction* (e.g., bit reset) when multiple computational paths converge, and a more complicated *prediction* along potentially diverging forward-looking computational paths. This distinction is addressed in the following section.

Connection to Landauer's limit. Turning our attention to quasistatic and not necessarily reversible processes, we note that in these processes $Z_1 \approx Z_2$, and under this approximation, equation (4) still holds without the additive constant $\log_2 \frac{Z_1}{Z_2}$. Furthermore, under constant temperature, the external entropy production is $\Delta S_{ext} = \int dq_{ext}/T = \Delta q_{ext}/T$, where q_{ext} represents the heat flow to the system from the exterior in the context of the source Y. Hence,

$$t_{Y \to X}(n+1) = -\frac{1}{k \log 2} \frac{\Delta q_{ext}}{T}, \tag{5}$$

In other words, for irreversible but quasistatic processes, local transfer entropy is proportional to the heat received or dissipated by the system from/to the exterior.

Thus, we observe that Landauer's limit $kT \log 2$ that associated a minimum entropy with a single bit of information is applicable here as well. In particular, for quasistatic processes, using (5), we obtain an equality that includes the classical Landauer's limit:

$$\Delta q_{ext} = -(kT \log 2) \, t_{Y \to X}(n+1). \tag{6}$$

Both of the considered examples, the Joule expansion of a one molecule gas and the Szilárd engine's compression (resetting one bit), can be interpreted using this equation. For the Joule expansion, $\Delta q_{ext} = 0$ due to thermal isolation, and there is a zero transfer entropy $t_{Y \to X} = 0$ due to independence of X and Y, so either side of the equation is trivially zero. During the bit reset by the Szilárd engine compression, heat is dissipated, yielding $\Delta q_{ext} = -kT \log 2$, while the transfer entropy $t_{Y \to X} = 1$, again in agreement with equation (6).

Landauer inequalities for non-equilibrium dynamics. Depending on the processes, heat transfer can occur at different temperatures, and, in general, $\int dq_{ext}/T \neq \Delta q_{ext}/T$. Neverthe-

less, under some stronger assumptions outlined in [28], the conditional entropies can be related to the heat transferred in the transition, per temperature, even when temperatures are varying. In a general non-equilibrium case, we may consider two cases: (i) when the system dissipates heat, transfer entropy is positive, and $Z_1 \leq Z_2$:

$$t_{Y \to X}(n+1) \leq -\frac{1}{k \log 2} \int \frac{dq_{ext}}{T}, \tag{7}$$

and (ii) when the system absorbs heat, transfer entropy is negative, and $Z_1 \geq Z_2$:

$$t_{Y \to X}(n+1) \geq -\frac{1}{k \log 2} \int \frac{dq_{ext}}{T}, \tag{8}$$

In the first case (a cooling system with positive transfer entropy), the negative $\int dq_{ext}/T$ is bounded from above by negative $-k \log 2 \, t_{Y \to X}(n+1)$, while in the second case (a heating system with negative transfer entropy), the positive $\int dq_{ext}/T$ is bounded from below by positive $-k \log 2 \, t_{Y \to X}(n+1)$. Generally, for absolute values,

$$\left| \int \frac{dq_{ext}}{T} \right| \geq k \log 2 \ |t_{Y \to X}(n+1)|. \tag{9}$$

For isothermal processes, this reduces to

$$|\Delta q_{ext}| \geq kT \log 2 \ |t_{Y \to X}(n+1)|. \tag{10}$$

For non-isothermal processes, a linear relationship between the transferred heat and temperature breaks down. For example, transfer entropy of a cycling system interacting with one hot (T_h) and one cold (T_c) thermal reservoirs, and exchanging with the surroundings the net heat $\langle q_{ext} \rangle$, can be bounded either as

$$0 \leq t_{Y \to X}(n+1) \leq -\frac{1}{k \log 2} \left(\frac{1}{T_c} - \frac{1}{T_h} \right) \langle q_{ext} \rangle = -\frac{\beta_c - \beta_h}{\log 2} \langle q_{ext} \rangle, \tag{11}$$

when $Z_1 \leq Z_2$, or

$$0 \leq -t_{Y \to X}(n+1) \leq \frac{1}{k \log 2} \left(\frac{1}{T_c} - \frac{1}{T_h} \right) \langle q_{ext} \rangle = \frac{\beta_c - \beta_h}{\log 2} \langle q_{ext} \rangle, \tag{12}$$

when $Z_1 \geq Z_2$.

Hence, we obtain an inequality involving a modified Landauer's limit:

$$|\langle q_{ext} \rangle| \geq \frac{\log 2}{\beta_c - \beta_h} |t_{Y \to X}(n+1)|. \tag{13}$$

The expressions (6) and (13) essentially set the "conversion rate" between transfer entropy and the dissipated/received heat. Specifically, when transfer entropy is increased within

the system by one bit, $t_{Y \to X}(n + 1) = 1$, the dissipated heat must be equal to (or larger than the modified) Landauer's limit. The obtained inequality is in agreement with the generalised Clausius statement considered by [29] in the context of information reservoirs and memory writing. The modified version of Landauer's principle offered in [29] appears to have a slight error in the concluding equation [50] where the corresponding $(\beta_{cold} - \beta_{hot})$ term is in the numerator rather than the denominator.

Intuitively, when the system is cooled by losing heat equivalent to Landauer's limit, the predictability about the system cannot be increased by more than one bit. This interpretation is non-trivial because Landauer's limit specifies the amount of heat needed to reset one bit of information (limiting *retrodiction*), i.e., information is destroyed because multiple computational paths converge, while when dealing with local transfer entropy one considers *prediction* along forward-looking computational paths which may diverge. Thus, the suggested interpretation generalises Landauer's limit to bi-directional information dynamics for quasistatic processes, and generalises the modified Landauer's limit to bi-directional information dynamics for non-equilibrium processes (subject to the aforementioned additional assumptions specified in [28]).

Connection to Bremermann's limit. In this section we analyze how fast a physical computing device can perform a logical operation, by connecting this problem to the dynamics of predictability captured by local transfer entropy. Margolus and Levitin [30] specified this question through computational *speed*: the maximum number of distinct states that the system can pass through, per unit of time, pointing out that, for a classical computer, this would correspond to the maximum number of operations per second. It is well-known that this quantity is limited (*Bremermann's limit* [31]), and is immediately connected to how much energy is available for information processing, e.g., for switching between distinct states. As pointed out by Margolus and Levitin [30], the rate at which a system can oscillate between two distinct states is $\nu_\perp \le \frac{4E}{h}$, where h is the Planck constant, and E is fixed average energy, assuming zero of energy at the ground state. For a quantum system, where distinct states are orthogonal, "the average energy of a macroscopic system is equal to the maximum number of orthogonal states that the system can pass through per unit of time" [30]. The limit is smaller when a sequence of oscillations is considered: $\nu_\perp \le \frac{2E}{h}$ for a long evolution through orthogonal states.

The work by Margolus and Levitin [30] strengthened a series of previous results which related the rate of information processing to the standard deviation of the energy: $\nu_\perp \le \frac{4\delta E}{h}$ (cf. [32]). While these previous results specified that a quantum state with *spread in energy δE* takes time at least $\Delta \tau = \frac{1}{\nu} = \frac{h}{4\delta E}$ to evolve to an orthogonal state, the result of Margolus and Levitin bounds the minimum time via the average energy E: $\Delta \tau \ge \frac{h}{4E}$, rather than via the spread in energy δE which can be arbitrarily large for fixed E.

Importantly, Margolus and Levitin [30] argued that these bounds are achievable for an ordinary macroscopic system (exemplified, for instance, with a lattice gas): "adding energy increases the maximum rate at which such a system can pass through a sequence of mutually orthogonal states by a proportionate amount" [30].

7

Following Lloyd [33] we interpret Bremermann's limit as the maximum number of logical operations that can be performed per second, i.e. $\nu = \frac{1}{\Delta\tau} = \frac{2E}{\pi\hbar}$, where $\hbar$ is the reduced Planck constant, and E is the energy of the system.

We now consider a transition during which the system dissipates energy, $\Delta E < 0$, noting that the notation ΔE for the energy change should not be confused with the spread in energy δE. Then we define the change in the *maximum* number of logical operations per second, i.e. the *computational deceleration*, as $\Delta\nu = \frac{2\Delta E}{\pi\hbar}$. This quantity accounts for how much the frequency of computation ν is reduced when the system loses some energy. This expression can also be written as $h\,\Delta\nu = 4\Delta E$, relating the energy change to the change in frequency.

Then, generalizing (10) to $|\Delta E| \geq kT\log 2\,|t_{Y\to X}|$, where ΔE is the heat energy which left the system X and entered the exterior Y during an isothermal process, we can specify a lower bound on the computational deceleration needed to increase predictability about the transition by $t_{Y\to X}$, at a given temperature:

$$\Delta\nu \geq \frac{2\log 2}{\pi\hbar}kT\,|t_{Y\to X}|\,. \tag{14}$$

That is, the expression (14) sets the *minimum* computational deceleration needed to increase predictability. For example, considering the Szilárd engine's compression resetting one bit, we note that the computational deceleration within such a device, needed to increase predictability about the transition by one bit, is at least $\frac{2\log 2}{\pi\hbar}kT$, or $\frac{4}{h}kT\log 2$. In other words, the product $h\,\Delta\nu$ is bounded from below by the energy equal to four times the Landauer's limit, that is, $h\,\Delta\nu \geq 4kT\log 2$.

Connection to the Bekenstein bound. Another important limit is the *Bekenstein bound*: an upper limit $\hat{I}$ on the information that can be contained within a given finite region of space which has a finite amount of energy E [34]:

$$\hat{I} = \frac{2\pi RE}{\hbar c\log 2}, \tag{15}$$

where R is the radius of a sphere that can enclose the given system, and c is the speed of light.

While Bremermann's limit constrains the rate of computation, the Bekenstein bound restricts an information-storing capacity [35]. It applies to physical systems of finite size with limited total energy and limited entropy, specifying the maximal extent of memory with which a computing device of finite size can operate.

Again considering a transition during which the system dissipates energy, $\Delta E < 0$, we define the change in the maximum information $\hat{I}$ required to describe the system, i.e., the *information loss*, as $\Delta\hat{I} = \frac{2\pi R\Delta E}{\hbar c\log 2}$.

Then the predictability during an isothermal state transition within this region, reflected

in the transfer entropy $t_{Y \to X}$ from a source Y, is limited by the information loss $\Delta \hat{I}$ associated with the energy change ΔE during the transition. Specifically, using $|\Delta E| \geq kT \log 2 \, |t_{Y \to X}|$ for any source Y, we obtain:

$$\Delta \hat{I} \geq \frac{2\pi R}{\hbar c} kT \, |t_{Y \to X}| \,. \tag{16}$$

While predictability about the system is increased by one bit, at a given temperature, i.e., $t_{Y \to X}(n+1) = 1$, there is a loss of at least $\frac{2\pi R}{\hbar c} kT$ from the maximum information contained within (or describing) the system.

Let us revisit the Szilárd engine compression during a bit reset operation, this time within a spherical container of fixed radius R which dissipates heat energy to the exterior. Such a compressed system needs less information to describe it than an uncompressed system, and the corresponding information loss about the compressed system $\Delta \hat{I}$ is bounded from below by Landauer's limit scaled with $\frac{2\pi R}{\hbar c \log 2}$.

It is important to realise that while the system dissipates energy and loses information/entropy, the increased predictability is about the transition. Therefore, this increased predictability reflects the change of information rather than the amount of information in the final configuration. Hence, the expressions (14) and (16) set *transient* limits of computation, for computational deceleration and information loss, respectively.

Discussion

The obtained relationships (6), (9), (13), (14) and (16) specify different *transient* computational limits bounding increases of predictability about a system during a *transition*, represented via transfer entropy. These relations explicitly identify constraints on transfer entropy in small scale physical systems, operating on the scale of the thermal energy kT, and being essentially 'quantized' by Landauer's limit. These constraints express increases of predictability via dissipated heat/energy; set the minimum computational deceleration needed to increase predictability; and offset the loss in the maximum information contained within a physical system by the predictability gained during a transition. Unlike classical Bremermann's limit and the Bekenstein bound which set the maximum computational speed and information, these inequalities specify the lower bounds, showing that in order to achieve a gain in predictability, the transient computational dynamics such as deceleration (or information loss) need to operate faster (or contain more) than these limits. Understanding these limits and their implications is becoming critical as computing circuitry is rapidly approaching these regimes [36].

Finally, we point out an important relationship between an increase of predictability about the system (higher local transfer entropy) and *negentropy*: the entropy that the system exports (dissipates) to keep its own entropy low [37]. That is, the expression (4) may indicate the general applicability to guided self-organisation in various artificial life scenarios, where one would expect that maximising transfer entropy corresponds to maximising negentropy. It is

9

known that negentropy $\Delta S_{ext} = -\Phi$, where Φ is the Massieu-Planck thermodynamic potential (free entropy). It was shown that maximising Φ is related to stability in several molecular biology contexts (e.g., protein stability [38]), and so the suggested thermodynamic interpretation associates such increases in stability with increases in transfer entropy to the system. One may also argue that the increase of stability in biological systems due to a free entropy change (measured in bits) is also scaled, and in some cases 'quantised', by Landauer's limit.

Methods

Preliminaries. Formally, consider a time-series process X of the (potentially multivariate) random variables $\{\ldots X_{n-1}, X_n, X_{n+1} \ldots\}$ with process realizations $\{\ldots x_{n-1}, x_n, x_{n+1} \ldots\}$ for countable time indices n. The underlying *state* of the process X is described by a time series of vectors $\{\ldots \mathbf{X_{n-1}}, \mathbf{X_n}, \mathbf{X_{n+1}} \ldots\}$ with realizations $\{\ldots \mathbf{x_{n-1}}, \mathbf{x_n}, \mathbf{x_{n+1}} \ldots\}$, where the multivariate realization $\mathbf{x_n}$ fully describes the state of the process at n, perhaps using vectors $\mathbf{x_n} = \{x_{n-k+1}, \ldots, x_{n-1}, x_n\}$ for a length k Markovian process, or for a thermodynamic process by including all relevant thermodynamic variables. If vectors $\{x_{n-k+1}, \ldots, x_{n-1}, x_n\}$ are interpreted as embedding vectors [39], as proxies of hidden states, one should in general be cautious to avoid false coupling detections [40].

The probability distribution function for observing a realization $\mathbf{x_n}$ is denoted by $p(\mathbf{x_n})$, while $p(\mathbf{x_{n+1}} \mid \mathbf{x_n}) = \frac{p(\mathbf{x_{n+1}}, \mathbf{x_n})}{p(\mathbf{x_n})}$, denotes conditional probability of observing realization $\mathbf{x_{n+1}}$ having observed $\mathbf{x_n}$ at the previous time step, where $p(\mathbf{x_{n+1}}, \mathbf{x_n})$ is the joint probability of the two realizations. These quantities are similarly defined for process Y, for corresponding time indices n.

Transfer entropy. The transfer entropy $T_{Y \to X}$, defined by (1), is a conditional mutual information [3] between $\mathbf{Y_n}$ and $\mathbf{X_{n+1}}$ given $\mathbf{X_n}$. Following Fano [41] we can quantify "the amount of information provided by the occurrence of the event represented by" $\mathbf{y_n}$ "about the occurrence of the event represented by" $\mathbf{x_{n+1}}$, conditioned on the occurrence of the event represented by $\mathbf{x_n}$. That is, we can quantify *local* or *point-wise transfer entropy* [10] in the same way Fano derived local or point-wise conditional mutual information [41]. This is a method applicable in general for different information-theoretic measures [42]: for example, local entropy or Shannon information content for an outcome x_n of process X is defined as $h(x_n) = -\log_2 p(x_n)$. The quantity $h(x_n)$ is simply the information content attributed to the specific symbol x_n, or the information required to predict or uniquely specify that specific value. Other local information-theoretic quantities may be computed as sums and differences of local entropies, e.g., $h(x_{n+1} \mid x_n) = h(x_{n+1}, x_n) - h(x_n)$, where $h(x_{n+1}, x_n)$ is the local joint entropy. In computing these quantities, the key step is to estimate the relevant probability distribution functions. This could be done using multiple realisations of the process, or accumulating observations in time while assuming stationarity of the process in time (and/or space for spatio-temporal processes).

As such, the local transfer entropy may be expressed as a difference between local conditional entropies:

$$t_{Y \to X}(n+1) = \log_2 \frac{p(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n})}{p(\mathbf{x_{n+1}} \mid \mathbf{x_n})} \tag{17}$$

$$= h(\mathbf{x_{n+1}} \mid \mathbf{x_n}) - h(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n}), \tag{18}$$

where local conditional entropies are defined as follows [3]:

$$h(\mathbf{x_{n+1}} \mid \mathbf{x_n}) = -\log_2 p(\mathbf{x_{n+1}} \mid \mathbf{x_n}), \tag{19}$$

$$h(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n}) = -\log_2 p(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n}). \tag{20}$$

Entropy definitions. The thermodynamic entropy was originally defined by Clausius as a state function S satisfying

$$S_B - S_A = \int_A^B \mathrm{d}q_{rev}/T, \tag{21}$$

where q_{rev} is the heat transferred to an equilibrium thermodynamic system during a reversible process from state A to state B. It can be interpreted, from the perspective of statistical mechanics, via the famous Boltzmann's equation $S = k \log W$, where W is the number of microstates corresponding to a given macrostate. Sometimes W is termed "thermodynamic probability" which is quite distinct from a mathematical probability bounded between zero and one. In general, W can be normalized to a probability $p = W/N$, where N is the number of possible microstates for all macrostates. This is not immediately needed as we shall consider (and later normalize) relative "thermodynamic probabilities".

At this stage, we recall a specialization of Boltzmann's principle by Einstein [43], for two states with entropies S and S_0 and "relative probability" W_r (the ratio of numbers W and W_0 that account for the numbers of microstates in the macrostates with S and S_0 respectively), given by: $S - S_0 = k \log W_r$. Here again the "relative probability" W_r is not bounded between zero and one. For instance, if the number of microstates in B (i.e., after the transition) is twice as many as those in A (as is the case, for instance, after a free adiabatic gas expansion; see the Joule expansion example described in the next section), then $W_r = 2$, and the resultant entropy change is $k \log 2$. Thus, the "relative probability" W_r depends on the states involved in the transition.

In general, the variation of entropy of a system $\Delta S = S - S_0$ is equal to the sum of the internal entropy production σ inside the system and the entropy change due to the interactions with the surroundings ΔS_{ext}:

$$\Delta S = \sigma + \Delta S_{ext}, \tag{22}$$

so that when the transition from the initial state S_0 to the final state S is irreversible, the entropy production $\sigma > 0$, while for reversible processes $\sigma = 0$.
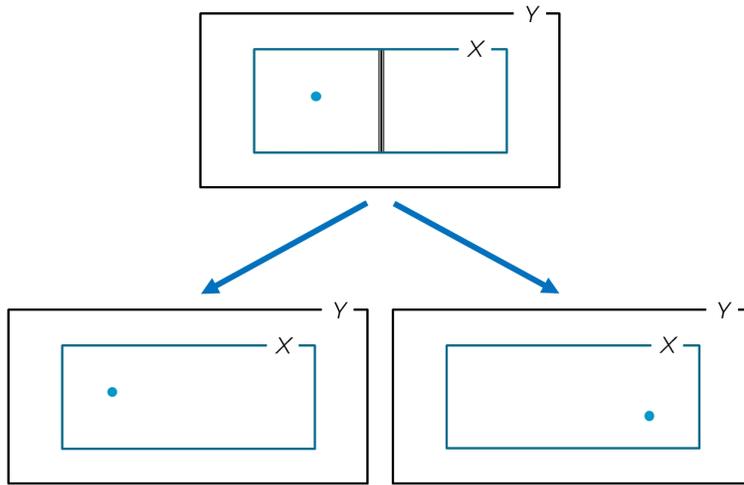
11

Figure 1: The Joule expansion of a one molecule gas. The container X is in thermal isolation from the surrounding exterior Y. As the partition is removed, the particle may be on the left or on the right.

Examples of coupled systems.

Joule expansion. The first example is the classical Joule expansion. A container X is thermally isolated from the heat bath Y, and there are no heat exchanges between X and Y. A partition separates two chambers of X, so that a volume of gas is kept in the left chamber of X, and the right chamber of X is evacuated. Then the partition between the two chambers is opened, and the gas fills the whole container, adiabatically. It is well-known that such (irreversible) doubling of the volume at constant temperature T increases the entropy of X, i.e. $\Delta S(x) = nk \log 2$, where n is the number of particles (cf. the Sackur-Tetrode equation [44]). The same increase of entropy results from a quasistatic irreversible variant of the Joule expansion. If there was just one particle in the container (a one molecule gas, as shown in Fig. 1), the entropy would still increase after the gas expansion, by $\Delta S(x) = k \log 2$, reflecting the uncertainty about the particle's position with respect to the container's chambers (left or right).

In this example, there are no interactions between the systems X (the container) and Y (the exterior). In other words, $\Delta S(x)_{ext} = 0$, and the entropy increase is due to the internal entropy production. Formally, $\Delta S(x) = \sigma(x)_y$, that is, the internal entropy produced by X is not affected at all by the system Y, because of the total thermal isolation between X and Y.

Szilárd engine compression. Now let us consider the second example with the same container X, surrounded by the exterior system Y, but now without a thermal isolation between the two, so that heat can be exchanged between X and Y. Two movable partitions can still divide the container into two separate chambers, as well as frictionlessly slide along the walls of the container to either side, the left or to the right. For clarity, we consider only one particle contained in the system X, and consider the dynamics of the partitions as part of the exterior

Y. This setup is (a part of) the Szilárd heat engine [45,46].

When a partition is in the middle of the container, the particle is located either in the left or the right chamber with equal probability, but eventually collisions between the particle and the partition force the latter to one side. It is easy to calculate that, as one partition isothermally moves to the side of the container (at temperature T), the maximum work extracted from the heat bath is $kT \log 2$. In order for the work to be extracted, one obviously needs to know on which side the particle was located initially. The most important aspects, in the context of our example, are how this coupled system modelled on the Szilárd engine realizes the logically irreversible operation of resetting a bit — the operation used to illustrate Landauer's principle — and what are the resultant entropy dynamics.

A similar physical implementation of this operation is, for instance, considered by Maroney [47], as shown in Fig. 2. If the particle is on the left side, then the physical state represents logical state zero, and if the particle is on the right side, it represents logical state one. The partition is removed from the middle of the container, allowing the particle to freely travel within the container. Then the partition is inserted into the far right-hand side of the container and is slowly (quasistatically) moved to its original position in the middle. This compression process maintains thermal contact between the container X and heat bath Y. Collisions with the particle exert a pressure on the partition, requiring work to be performed, and the energy from the work is transferred to heat in the heat bath Y, amounting to at least $kT \log 2$. Such resetting of a bit to zero, which converts at least $kT \log 2$ of work into heat, is a typical example of Landauer's principle: any logically irreversible transformation of classical information is necessarily accompanied by the dissipation of at least $kT \log 2$ of heat per reset bit. The entropy dynamics is the same for the Szilárd engine compression, where one partition is fixed at the edge and only the second one is moving.

It is easy to see that the physical result of such isothermal compression is a decrease in the thermodynamical entropy of the one molecule gas by $k \log 2$, accompanied by an increase in the entropy of the environment by (at least) the same amount. That is, $\Delta S(x) = -k \log 2$. At the same time, as the example shows, the heat is dissipated to the exterior system Y, i.e., compensated by the entropy change due to the interactions with the surroundings, $\Delta S(x)_{ext} = -k \log 2$, and hence, there is no internal entropy production, $\sigma(x)_y = 0$.

Assumptions and their illustration in the examples. In an attempt to provide a thermodynamic interpretation of transfer entropy two important assumptions are made [28], defining the range of applicability for such an interpretation. The first one relates the transition probability W_{r_1} of the system's reversible state change to the conditional probability $p(\mathbf{x_{n+1}} \mid \mathbf{x_n})$:

$$p(\mathbf{x_{n+1}} \mid \mathbf{x_n}) = \frac{1}{Z_1} W_{r_1} \tag{23}$$

where Z_1 is a normalization factor, and W_{r_1} is such that $S(\mathbf{x_{n+1}}) - S(\mathbf{x_n}) = k \log W_{r_1}$. The normalization factor is equal to the ratio between the total number of microstates in all possible
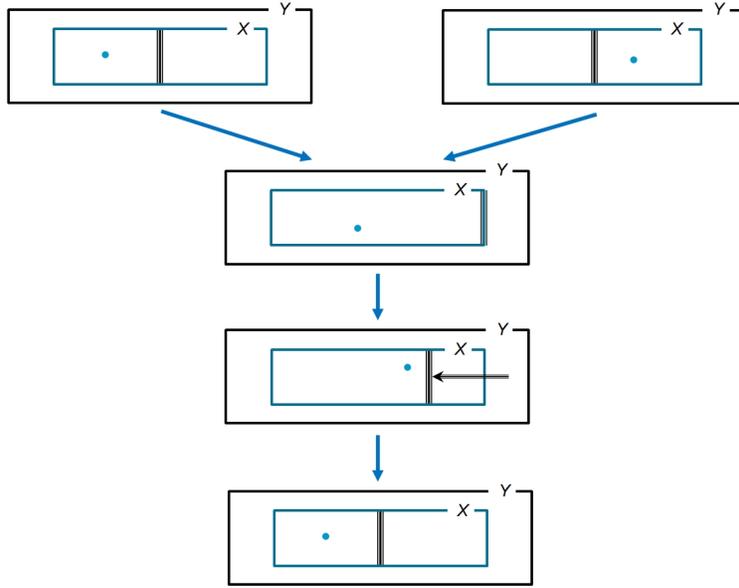
Figure 2: Resetting one bit by compression in the Szilárd engine-like device. The container X is in thermal contact with the surrounding exterior Y. As the partition is re-inserted and moved back to the middle, the particle returns to the left side.

macrostates at $n+1$ and the total number of microstates in all possible macrostates at n.

We note that the normalization is needed because "relative probability" W_{r_1} is not bounded between zero and one, while the conditional probability is a properly defined mathematical probability. For example, in the Joule expansion, the number of microstates in $\mathbf{x_{n+1}}$ is twice as many as those in the old state $\mathbf{x_n}$, making $W_{r_1} = 2$. In this simple example the state $\mathbf{x_{n+1}}$ is the only possible macrostate, and the normalization factor $Z_1 = 2$ as well, since the number of all microstates *after* the transition is twice the number of the microstates in the old state before the transition. Hence, the right-hand side of the assumption is unity, $\frac{1}{Z_1}W_{r_1} = 1$, concurring with the conditional probability $p(\mathbf{x_{n+1}} \mid \mathbf{x_n}) = 1$, as this transition is the only one possible. The entropy change is $S(\mathbf{x_{n+1}}) - S(\mathbf{x_n}) = k\log 2$, as expected.

On the contrary, for the Szilárd engine's compression example, the number of microstates in $\mathbf{x_n}$ *before* the transition is twice as many as those in the new state $\mathbf{x_{n+1}}$ after the transition (e.g., partition moves from the far right-hand side to the middle), and $W_{r_1} = 1/2$. Following Szilárd's original design, we allow for two possibly moving partitions: either from the right to the middle, or from the left to the middle, with both motions resulting in the same entropy dynamics. Hence, the total number of possible microstates in all the resultant macrostates is still 2: the particle is on the left, or on the right, respectively. Therefore, there is no overall reduction in the number of all possible microstates after the transition from $\mathbf{x_n}$ to $\mathbf{x_{n+1}}$, so that $Z_1 = 1$. Thus, $\frac{1}{Z_1}W_{r_1} = 1/2$, setting the conditional probability $p(\mathbf{x_{n+1}} \mid \mathbf{x_n}) = 1/2$ for the partition moving from the right. The entropy change is given by $S(\mathbf{x_{n+1}}) - S(\mathbf{x_n}) = k\log 1/2 = -k\log 2$.

The second assumption relates the transition probability W_{r_2} of the system's internal state change, in the context of the interactions with the external world represented in the state vector $\mathbf{y}$, to the conditional probability $p(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n})$. Specifically, the second assumption is set as follows:

$$p(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n}) = \frac{1}{Z_2}W_{r_2} \tag{24}$$

for some number W_{r_2}, such that $\sigma(x)_y = k \log W_{r_2}$, where $\sigma(x)_y$ is the system's internal entropy production in the context of $\mathbf{y}$. The normalization factor is again the ratio between the total numbers of possible microstates after and before the transition. In general, $Z_1 \neq Z_2$ because $\mathbf{y_n}$ may either increase or constrain the number of microstates in $\mathbf{x_{n+1}}$.

For the Joule expansion example, $W_{r_2} = W_{r_1} = 2$, $Z_2 = Z_1 = 2$, with identical resultant probabilities. The lack of difference is due to thermal isolation, and hence independence, between X and Y, yielding $\Delta S(x) = \sigma(x)_y = k \log 2$.

For the Szilárd engine's compression example, $Z_2 = Z_1 = 1$. There is no internal entropy production, $\sigma(x)_y = k \log W_{r_2} = 0$, entailing $W_{r_2} = 1$. Thus, our second assumption requires that the resultant conditional probability is set as $p(\mathbf{x_{n+1}} \mid \mathbf{x_n}, \mathbf{y_n}) = \frac{1}{Z_2}W_{r_2} = 1$. That is, once the direction of the partition's motion is selected (e.g., from the right to the middle), the transition is certain to reach the compressed outcome, in context of the external contribution from Y.

1. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).

2. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 and 623–656 (1948).

3. MacKay, D. J. *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2003).

4. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (John Wiley & Sons, 1991).

5. Baek, S. K., Jung, W.-S., Kwon, O. & Moon, H.-T. Transfer entropy analysis of the stock market (2005). ArXiv:physics/0509014v2.

6. Moniz, L. J., Cooch, E. G., Ellner, S. P., Nichols, J. D. & Nichols, J. M. Application of information theory methods to food web reconstruction. *Ecol. Model.* **208**, 145–158 (2007).

7. Chávez, M., Martinerie, J. & Le Van Quyen, M. Statistical assessment of nonlinear causality: application to epileptic EEG signals. *J. Neurosci. Methods* **124**, 113–128 (2003).

8. Wibral, M. *et al.* Transfer entropy in magnetoencephalographic data: quantifying information flow in cortical and cerebellar networks. *Prog. Biophys. Mol. Bio.* **105**, 80–97 (2011).

9. Pahle, J., Green, A. K., Dixon, C. J. & Kummer, U. Information transfer in signaling pathways: a study using coupled simulated and experimental data. *BMC Bioinformatics* **9**, 139 (2008).

10. Lizier, J. T., Prokopenko, M. & Zomaya, A. Y. Local information transfer as a spatiotemporal filter for complex systems. *Phys. Rev. E* **77**, 026110 (2008).

11. Lizier, J. T., Prokopenko, M. & Zomaya, A. Y. Information modification and particle collisions in distributed computation. *Chaos* **20**, 037109 (2010).

12. Lizier, J. T., Prokopenko, M. & Zomaya, A. Y. Local measures of information storage in complex distributed computation. *Inform. Sciences* **208**, 39–54 (2012).

13. Barnett, L. & Bossomaier, T. Transfer entropy as a log-likelihood ratio. *Phys. Rev. Lett.* **109**, 138105 (2012).

14. Wang, X. R., Miller, J. M., Lizier, J. T., Prokopenko, M. & Rossi, L. F. Quantifying and tracing information cascades in swarms. *PLoS ONE* **7**, e40084 (2012).

15. Lizier, J. T., Pritam, S. & Prokopenko, M. Information dynamics in small-world Boolean networks. *Artif. Life* **17**, 293–314 (2011).

16. Ceguerra, R. V., Lizier, J. T. & Zomaya, A. Y. Information storage and transfer in the synchronization process in locally-connected networks. In *2011 IEEE Symposium on Artificial Life (ALIFE), April 1315, 2011, Paris, France*, 54–61 (IEEE, New York, 2011).

17. Lungarella, M. & Sporns, O. Mapping information flow in sensorimotor networks. *PLoS Comput. Biol.* **2**, e144 (2006).

18. Lizier, J. T., Prokopenko, M. & Zomaya, A. Y. The information dynamics of phase transitions in random Boolean networks. In Bullock, S., Noble, J., Watson, R. & Bedau, M. A. (eds.) *Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems (ALife XI), Winchester, UK*, 374–381 (MIT Press, Cambridge, MA, 2008).

19. Barnett, L., Lizier, J. T., Harré, M., Seth, A. K. & Bossomaier, T. Information flow in a kinetic ising model peaks in the disordered phase. *Phys. Rev. Lett.* **111**, 177203 (2013).

20. Bennett, C. H. Notes on Landauer's principle, reversible computation, and Maxwell's Demon. *Stud. Hist. Phil. Sci. B* **34**, 501–510 (2003).

21. Piechocinska, B. Information erasure. *Phys. Rev. A* **61**, 062314 (2000).

22. Lloyd, S. *Programming the Universe* (Vintage Books, New York, 2006).

23. Parrondo, J. M. R., den Broeck, C. V. & Kawai, R. Entropy production and the arrow of time. *New J. Phys.* **11**, 073008 (2009).

24. Prokopenko, M., Lizier, J. T., Obst, O. & Wang, X. R. Relating Fisher information to order parameters. *Phys. Rev. E* **84**, 041116 (2011).

25. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191 (1961).

26. Goyal, P. Information physics – towards a new conception of physical reality. *Information* **3**, 567–594 (2012).

27. Landauer, R. Information is physical. *Phys. Today* **44**, 23–29 (1991).

28. Prokopenko, M., Lizier, J. T. & Price, D. C. On thermodynamic interpretation of transfer entropy. *Entropy* **15**, 524–543 (2013).

29. Deffner, S. & Jarzynski, C. Information processing and the second law of thermodynamics: an inclusive, Hamiltonian approach (2013). ArXiv:1308.5001v1.

30. Margolus, N. & Levitin, L. B. The maximum speed of dynamical evolution. *Phys. D* **120**, 188–195 (1998).

31. Bremermann, H. J. Minimum energy requirements of information transfer and computing. *Int. J. Theor. Phys.* **21**, 203–217 (1982).

32. Anandan, J. & Aharonov, Y. Geometry of quantum evolution. *Phys. Rev. Lett.* **65**, 1697–1700 (1990).

33. Lloyd, S. Ultimate physical limits to computation. *Nature* **406**, 1047–1054 (2000).

34. Bekenstein, J. D. Universal upper bound on the entropy-to-energy ratio for bounded systems. *Phys. Rev. D* **23**, 287–298 (1981).

35. Bekenstein, J. D. Entropy content and information flow in systems with limited energy. *Phys. Rev. D* **30**, 1669–1679 (1984).

36. Bérut, A. *et al.* Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* **483**, 187–189 (2012).

37. Schrödinger, E. *What is life? The Physical Aspect of the Living Cell.* (Cambridge University Press, 1944).

38. Schellman, J. A. Temperature, stability, and the hydrophobic interaction. *Biophys. J.* **73**, 2960–2964 (1997).

39. Takens, F. Detecting strange attractors in turbulence. In Rand, D. & Young, L.-S. (eds.) *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics, 366–381 (Springer, Berlin / Heidelberg, 1981).

40. Smirnov, D. A. Spurious causalities with transfer entropy. *Phys. Rev. E* **87**, 042917 (2013).

41. Fano, R. M. *The Transmission of Information: A Statistical Theory of Communication* (MIT Press, Cambridge, Massachussets, 1961).

42. Lizier, J. T. Measuring the dynamics of information processing on a local scale in time and space. In Wibral, M., Vicente, R. & Lizier, J. T. (eds.) *Directed Information Measures in Neuroscience*, Understanding Complex Systems, 161–193 (Springer Berlin Heidelberg, 2014).

43. Einstein, A. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Ann. Phys.* **322**, 132–148 (1905).

44. Huang, K. *Introduction to Statistical Physics*. A Chapman & Hall book (Taylor & Francis Group, 2009).

45. Szilárd, L. Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen (On the reduction of entropy in a thermodynamic system by the intervention of intelligent beings). *Zeitschrift für Physik* **53**, 840–856 (1929).

46. Magnasco, M. O. Szilard's heat engine. *Europhys. Lett.* **33**, 583–588 (1996).

47. Maroney, O. Information processing and thermodynamic entropy. In Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (2009), Fall 2009 edn, Date of access: 08/10/2013, URL = http://plato.stanford.edu/archives/fall2009/entries/information-entropy/.