6 Thermodynamic cost of measurement

In the two previous chapters, we have discussed the effects of information on the second law, in particular how information increases the capacity of a Maxwell demon to extract energy from a thermal bath. However, we have not yet addressed the more fundamental issue of how to restore the validity of the second law by exploring the thermodynamic cost of the demon's operations. This corresponds to the second task of the thermodynamics of information, as stated in the introduction and in chapter 1.

Now we discuss this issue by considering the physical nature of the demon, which can be a single degree of freedom that interacts with the original system or a complex device comprising a measurement apparatus, a processing unit to infer the feedback action from the measurement outcome, and some machinery to operate on the original system and implement that action.

To model such a variety of situations, we first distinguish between autonomous and non-autonomous or driven demons. In the first case, system and demon evolve without any external driving. To keep the global system out of equilibrium, it is necessary to couple it with several thermal baths and chemostats with different temperatures and/or chemical potentials. Then, the global system reaches a non-equilibrium steady state where one of the two subsystems, the demon, monitors the other one and affects its dynamics. We will explore, at the end of the chapter, the concept of information flow, which allows one to interpret along those lines a non-equilibrium bipartite system as a demon-system pair.

On the other hand, the original Szilárd engine consists of a sequence of steps that cannot be modeled as a non-equilibrium steady state. An external agent is necessary to initiate and terminate the interaction that correlates the system and the demon in the measurement, as well as to implement the feedback protocol. This external agent can be seen as a part of the demon. However, the corresponding driving cannot use explicitly the information gathered in the measurement; it must be a "blind" protocol that only activates the demon's machinery. Notice also that this blind protocol can be converted into an autonomous Maxwell demon replacing the driven parameters by degrees of freedom with a big mass and a non-zero initial velocity, as explained in the box *From driven to autonomous systems*, section 3.3.1.

In non-autonomous Maxwell demons, such as the Szilárd engine and other feedback motors, the different stages —namely: measurement, feedback protocol, and memory resetting— can be analyzed separately. We first study in detail the energetics of an ideal classical measurement in section 6.1. For this purpose, we use non-equilibrium free energy and the second law (3.53). Then, we combine in section 6.2 this analysis and the results of previous chapters on the minimal work in feedback processes and in the manipulation of information devices to derive bounds for the minimal work needed to carry out the different stages of the Szilárd engine and other generic feedback processes. Our analysis will reveal that the correlation between the system and the demon, as measured by mutual information, is in fact a form of free energy and that the Szilárd engine can be interpreted as the creation and exploitation of those correlations. This interpretation can be extended to autonomous bipartite systems evolving in continuous time, with the introduction of information flows, a concept that we discuss in section 6.3 and apply to a specific example in section 6.4.

6.1 Classical ideal measurements

In this section, we analyze the energetics of an ideal measurement, where an observer or demon interacts with a system and the probabilistic informational state M of the former correlates with the microscopic state X of the latter. We model the demon as an information device like the ones introduced in section 4.2. Accordingly, the phase space of the demon Γ is partitioned into M regions Γ_m, which define the function $m(y) = m$ if $y \in \Gamma_m$, y being a microscopic state of the demon.

The measurement is characterized by the conditional probability distribution $p_{M|X}(m|x)$, relating the microscopic state of the system x and the measurement outcome m. For instance, in an error-free measurement, the outcome is a deterministic function $\mu(x)$ of the micro-state x of the system, and $p_{M|X}(m|x) = \delta_{m,\mu(x)}$. Other situations where the measurement is affected by noise are well described by $p_{M|X}(m|x)$, as illustrated in exercises 1.2, 2.6, and 4.1.

We define an ideal classical measurement as an interaction that correlates the demon and the system while keeping the latter unaltered. This is the main characteristic of an ideal classical measurement, but it is useful to add extra assumptions that simplify the analysis. To formulate in a precise form these assumptions, we denote by X_0 and Y_0 the random variables representing respectively the micro-state of the system and the demon before the measurement and by X_1 and Y_1 the corresponding micro-states after the measurement. An ideal classical measurement is characterized by the following properties:

i) System and demon are independent before the measurement:

$$\rho_{X_0Y_0}(x, y) = \rho_{X_0}(x)\rho_{Y_0}(y). \tag{6.1}$$

ii) System and demon do not interact before and after the measurement, i.e., the Hamiltonian of the global system before and after the measurement reads, respectively:

$$\mathcal{H}_{\rm tot}(x,y) = \mathcal{H}_{\rm sys}(x) + \mathcal{H}_{\rm dem}(y) \tag{6.2}$$
$$\mathcal{H}'_{\rm tot}(x,y) = \mathcal{H}'_{\rm sys}(x) + \mathcal{H}'_{\rm dem}(y) \tag{6.3}$$

iii) The system is not altered by the measurement:

$$\rho_{X_1}(X) = \rho_{X_0}(X) \qquad \text{and} \qquad \mathcal{H}_{\rm sys}(x) = \mathcal{H}'_{\rm sys}(x) \tag{6.4}$$

iv) The micro-state of the demon Y_1 depends on X_1 only through the outcome M, i.e.,

$$\rho_{Y_1|MX_1}(y|m,x) = \rho_{Y_1|M}(y|m). \tag{6.5}$$

This condition implies that three variables, X, Y, and M, form a Markov chain $X \leftrightarrow M \leftrightarrow Y$ (Cover and Thomas, 2006).

v) The micro-state of the demon Y_1 determines the informational state M, i.e.,

$$M = m(Y_1) \Rightarrow \rho_{Y_1|M}(y|m) = \begin{cases} \dfrac{\rho_{Y_1}(y)}{p_M(m)} & \text{if } m = m(y) \\ 0 & \text{otherwise.} \end{cases} \tag{6.6}$$

These properties determine the joint probability density for the micro-states after the measurement. The joint distribution can be written as

$$\begin{aligned} \rho_{X_1Y_1}(x,y) &= \sum_m \rho_{MX_1}(m,x)\rho_{Y_1|MX_1}(y|m,x) \\ &= \sum_m \rho_{X_1}(x)p_{M|X_1}(m|x)\rho_{Y_1|MX_1}(y|m,x). \end{aligned} \tag{6.7}$$

In this expression we can directly use (6.4) and (6.5), yielding

$$\rho_{X_1Y_1}(x,y) = \sum_m \rho_{X_0}(x)p_{M|X_1}(m|x)\rho_{Y_1|M}(y|m). \tag{6.8}$$

We can now compute the non-equilibrium free energy of the global system before and after the measurement. Since system and demon are initially uncorrelated, Eq. (6.1), and do not interact, Eq. (6.2), the free energy is additive:

$$\mathcal{F}(X_0, Y_0) = \mathcal{F}(X_0) + \mathcal{F}(Y_0). \tag{6.9}$$

After the measurement, the average energy is again additive, according to Eq. (6.3), and the joint Shannon information can be expressed in terms of the individual Shannon entropies and the mutual information using Eq. (2.18):

$$\begin{aligned} \mathcal{F}(X_1,Y_1) &= \langle \mathcal{H}_{\rm sys}(X_1)\rangle + \langle \mathcal{H}'_{\rm dem}(Y_1)\rangle - TS(X_1,Y_1) \\ &= \mathcal{F}(X_0) + \mathcal{F}(Y_1) + TI(X_1;Y_1) \\ &= \mathcal{F}(X_0) + \mathcal{F}(Y_1) + TI(X_1;M). \end{aligned} \tag{6.10}$$

Here we have used that $I(X_1; M) = I(X_1; Y_1)$, a relationship that follows from Eq. (6.5) and Eq. (6.6) (see exercise 6.2), and express the fact that the demon only gathers information of the state of the system X_1 through the measurement outcome M.

Finally, we can apply the second law for processes involving non-equilibrium states, Eq. (3.53), to find the minimal work needed to complete the measurement:

$$W_{\rm meas} = \Delta\mathcal{F} = \Delta\mathcal{F}_Y + TI(X_1; M) \tag{6.11}$$

where $\Delta\mathcal{F}_Y \equiv \mathcal{F}(Y_1) - \mathcal{F}(Y_0)$. Since $I(X_1; M)$ is positive, we see that measuring or, more generally, creating correlations between two systems, increases the free energy and, if not counterbalanced by $\Delta\mathcal{F}_Y$, needs work and heat dissipation to be completed. Recall that these work is done by the external agent that initiates and terminates the interaction between system and the memory of the demon.

6.2 The Szilárd engine revisited

We can now analyze the Szilárd engine considering the physical nature of both the single-particle gas and the demon. The Szilárd cycle consists of a measurement and the corresponding feedback protocol, in which a maximum work $TI(X_1; M)$ can be extracted. As first noticed by Bennett (Bennett, 1982), to check the validity of the second law we have to complete the cycle by resetting the demon to its initial state. The three stages of the whole cyclic process are depicted in figure 6.1. The energetics of the measurement has been analyzed in the previous section. Here we continue by analyzing the second subprocess: the feedback protocol. The final non-equilibrium free energy after this protocol reads

$$\mathcal{F}(X_2, Y_2) = \mathcal{F}(X_2) + \mathcal{F}(Y_2) + TI(X_2; Y_2) \tag{6.12}$$

where X_2 and Y_2 denote respectively the micro-states of the system and the demon after the feedback protocol. Since this protocol is a cycle for the system in the Szilárd engine, its marginal probability density and Hamiltonian are equal at the beginning and at the end; hence $\mathcal{F}(X_2) = \mathcal{F}(X_1) = \mathcal{F}(X_0)$. We also assume that the state of the demon is not affected by the feedback protocol, $\mathcal{F}(Y_2) = \mathcal{F}(Y_1)$. Then, the minimal work to complete it is

$$W_{\rm fb} = \mathcal{F}(X_2, Y_2) - \mathcal{F}(X_1, Y_1) = T\left[I(X_2; Y_2) - I(X_1; M)\right]. \tag{6.13}$$

As explained in sections 4.1 and 5.3, in an optimal feedback protocol, like the original Szilárd cycle, the work is $W_{\rm fb} = -TI(X_1; M)$. Then, we conclude that such optimal protocol is not only reversible, as discussed in section 5.3, but should also destroy all correlations between the system and the demon to satisfy $I(X_2; Y_2) = 0$.

Here a clarification on the origin of this work is in order. In the first chapters, the demon is visualized as a conscious being or complicated device that is able to measure and manipulate the system using

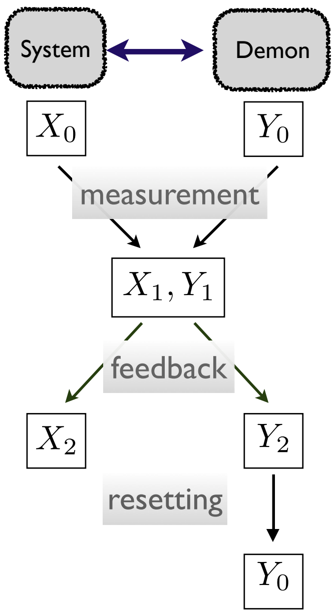


Fig. 6.1 Schematic representation of a feedback engine.

the information of the measurement. This manipulation is accompanied by the feedback work, bounded by $-TI(X_1; M)$ in Eq. (6.13). On the other hand, in the analysis of the measurement carried out in the previous section, we were forced to include an external agent to initiate the interaction needed to correlate the demon's memory with the state of the system. This external agent could be a part of the whole device that acts as a demon in the Szilárd engine. We then distinguish the memory and the operating machinery that constitutes the demon. This external agent can manipulate both the demon's memory and the system, but it cannot use the information stored in the memory, i.e., it must implement non-feedback protocols. This is why we use the second law for non-feedback processes Eq. (3.53), and not (4.8), to bound the work done by the external agent in Eq. (6.13). Finally, let us recall that this non-feedback driving could be converted into an autonomous Maxwell demon by introducing a degree of freedom with a large mass, as explained in the box *From driven to autonomous systems*, section 3.3.1.

Finally, the minimal work to reset the demon to its initial state Y_0 is

$$W_{\text{reset}} = \mathcal{F}(Y_0) - \mathcal{F}(Y_2) = -\Delta\mathcal{F}_Y. \tag{6.14}$$

The total minimal work $W_{\text{meas}} + W_{\text{fb}} + W_{\text{reset}}$, where the terms are given respectively by (6.11), (6.13), and (6.14), is zero if all the correlations are exploited, i.e., if $I(X_2; Y_2) = 0$. We then recover the standard second law for non-feedback cycles. Therefore, by incorporating the physical nature of the demon, we are able to restore the validity of the second law of thermodynamics for isothermal processes. Notice that the cost that compensates the extracted work $W_{\text{meas}} + W_{\text{reset}} = -W_{\text{fb}}$ can be distributed between the measurement and the resetting of the demon's memory, depending on the characteristics of the demon or, more precisely, on the free energy difference $\Delta\mathcal{F}_Y$. For instance, if the demon is a symmetric memory in local equilibrium with two informational states, L (left) and R (right), and its initial probabilistic informational state is given by $p_L = 0$ and $p_R = 1$, then $\Delta\mathcal{F}_Y = -kT\ln 2$. In this case, the measurement can be carried out at zero energy cost, but resetting the demon's memory is a Landauer's overwriting that requires a work $kT\ln 2$ and dissipates the same amount of heat to the thermal bath. On the other hand, if the initial state is $p_L = p_R = 1/2$, we have $\Delta\mathcal{F}_Y = 0$. Resetting is not necessary but the minimal work to measure is $kT\ln 2$.

Our analysis shows that the Szilárd engine or any generic feedback process can be interpreted as the creation of correlations between the system and the demon in the measurement, with the corresponding thermodynamic cost, followed by the exploitation of the free energy stored in those correlations.

This interpretation can be extended to systems that evolve in continuous time as $X(t)$, $Y(t)$, creating and destroying correlations. In fact, the analysis is easier and more general if we use the Shannon entropy $S(X(t), Y(t))$ of the global system. If we differentiate with respect to time the equation relating the mutual information and the joint entropy, we get

$$\dot{S}(X,Y) = \dot{S}(X) + \dot{S}(Y) - \dot{I}(X;Y), \tag{6.15}$$

then the entropy production rate can be written as

$$\dot{S}_{\text{prod}} = \dot{S}(X) + \dot{S}(Y) - \dot{I}(X;Y) + \dot{S}_{\text{env}} \geq 0 \tag{6.16}$$

where $\dot{S}_{\text{env}}$ is the change of the entropy of the environment per unit of time. In the case of a single thermal bath at temperature T, this change is $\dot{S}_{\text{env}} = -\dot{Q}/T$, and the previous equation is equivalent to

$$\begin{aligned} T\dot{S}_{\text{prod}} &= T\dot{S}(X) + T\dot{S}(Y) - T\dot{I}(X;Y) - \dot{Q} \\ &= \dot{W} - \dot{\mathcal{F}}(X) - \dot{\mathcal{F}}(Y) - T\dot{I}(X;Y) - \langle\dot{\mathcal{H}}_{\text{int}}\rangle \end{aligned} \tag{6.17}$$

where $\mathcal{H}_{\text{int}}(X,Y)$ is the interaction energy between the two systems. Here we have used the first law: $\langle\dot{\mathcal{H}}_X + \dot{\mathcal{H}}_Y + \dot{\mathcal{H}}_{\text{int}}\rangle = \dot{W} + \dot{Q}$. If this interaction energy vanishes at the beginning and at the end of a process, the integral over time of (6.17) yields

$$TS_{\text{prod}} = W - \Delta\mathcal{F}(X) - \Delta\mathcal{F}(Y) - T\Delta I(X;Y) \geq 0. \tag{6.18}$$

This expression includes as special cases the bounds found for the work needed to measure, complete a feedback cycle, and reset the memory, i.e., Eqs. (6.11), (6.13), and (6.14). Notice, however, that (6.18) is a special case of (6.16). The latter can be applied to more complicated cases where the global system $(X(t), Y(t))$ evolves in contact with several thermal baths with different temperatures and/or chemostats or particle reservoirs with different chemical potentials.

6.3 Information flows

Eq. (6.16) provides a simple picture of the Szilárd engine and generic non-autonomous Maxwell demons. For autonomous systems, on the other hand, this approach is not useful, because, if the system reaches a stationary state, even far from equilibrium, the mutual information $I(X(t);Y(t))$ between any pair of variables that describe parts of the system and the corresponding Shannon entropies, $S(X(t))$ and $S(Y(t))$, are constant. Hence, all the terms on the right-hand side of Eq. (6.16) vanish, except $\dot{S}_{\text{env}}$. An interesting attempt to overcome this issue is to decompose the change in mutual information

$$\dot{I}(X(t);Y(t)) = \dot{I}^X(t) + \dot{I}^Y(t) \tag{6.19}$$

into information flows, which are defined as

$$\dot{I}^X(t) = \left.\frac{d}{dt'}\right|_{t'=t} I(X(t');Y(t)) \tag{6.20}$$

$$\dot{I}^Y(t) = \left.\frac{d}{dt'}\right|_{t'=t} I(X(t);Y(t')). \tag{6.21}$$

This decomposition is only possible if $X(t+\Delta t)$ and $Y(t+\Delta t)$, conditioned to $X(t)$ and $Y(t)$, are independent up to order $(\Delta t)^2$, i.e.

$$\begin{aligned}\rho(x',y';t+\Delta t|x,y;t) &= \rho(x';t+\Delta t|x,y;t)\rho(y';t+\Delta t|x,y;t)\\ &+ o\left(\Delta t^2\right). \end{aligned} \quad (6.22)$$

Here, we have employed the abbreviated notation for the conditional distribution: $\rho_{X(t')Y(t')|X(t)Y(t)}(x',y'|x,y) = \rho(x',y';t'|x,y;t)$, and a similar notation for the marginal distributions. This condition is equivalent to [1] (see exercise 6.3):

$$\left.\frac{\partial}{\partial t'}\right|_{t'=t}\left[\rho_{X(t')Y(t)}(x,y)+\rho_{X(t)Y(t')}(x,y)\right] = \frac{d}{dt}\rho_{X(t)Y(t)}(x,y). \quad (6.23)$$

The condition is fulfilled by continuous Markov processes that are solutions of stochastic differential equations with uncorrelated noises, as well as by bipartite Markov chains $(X(t),Y(t))$, defined as those in which transitions that simultaneously change the state of X and Y are forbidden (see exercise 6.5). We study this case in detail below following (Horowitz and Esposito, 2014)

For a generic bipartite system, the first term in the decomposition (6.19), $\dot{I}^X(t)$, is the increase of the mutual information between the two systems *due solely to the evolution of* X. If this flow is positive, then the evolution of $X(t)$, keeping $Y(t)$ fixed, increases the mutual information. We can interpret this behavior as system X measuring system Y. On the other hand, if $\dot{I}^X(t)$ is negative, the evolution of X decreases the mutual information and this decrease can be used to extract energy from reservoirs. In this case, the evolution of system X is exploiting the free energy stored in the correlations, like the working substance of a feedback engine. In the stationary regime $\dot{I}(X(t);Y(t)) = 0$ and $\dot{I}^X(t) = -\dot{I}^Y(t)$; hence, according to the sign of the information flow, we can identify one of the two systems as the sensor or demon, and the other one as the working substance on which the feedback action is applied. For instance, if $\dot{I}^X(t)$ is positive, then system X acts as the demon and system Y as the working substance.

6.3.1 Bipartite Markov chains

To further clarify the interpretation of the information flow and show its utility, let us focus on a bipartite Markov chain $(X(t),Y(t))$. The reader can find a brief review of Markov chains in Appendix D. In the case of a bipartite system, the simultaneous change of both random variables is forbidden; therefore, the transition rate from state (x,y) to state (x',y') is of the form:

$$\Gamma[(x,y)\to(x',y')] = 0 \qquad \text{if } x\neq x' \text{ and } y\neq y', \quad (6.24)$$

[1]This expression is valid for any differentiable function $f(t,t')$. However, the joint probability distribution $\rho_{X(t')Y(t)}(x,y)$ is not differentiable on the line $t'=t$. For instance, for stationary Markov processes $\rho_{X(t')Y(t)}(x,y)$ is a function of $|t-t'|$ and $\min\{t,t'\}$. It is therefore continuous in the whole plane (t,t') but not differentiable on the line $t=t'$, even for bipartite systems.

and can be written as the rate of a jump in X, given a value of Y:

$$\Gamma^{y}_{x\to x'} \equiv \Gamma[(x,y)\to(x',y)] \tag{6.25}$$

and vice versa:

$$\Gamma^{x}_{y\to y'} \equiv \Gamma[(x,y)\to(x,y')]. \tag{6.26}$$

In a bipartite system, one can easily split the global evolution into two terms, each attributed to the respective system. The contribution due to the evolution of system X is

$$\left.\frac{\partial}{\partial t'}\right|_{t'=t} \rho_{X(t')Y(t)}(x,y) = \sum_{x'} J^{y}_{x'\to x}(t) \tag{6.27}$$

where

$$J^{y}_{x'\to x}(t) = \Gamma^{y}_{x'\to x}\rho_{X(t)Y(t)}(x',y) - \Gamma^{y}_{x\to x'}\rho_{X(t)Y(t)}(x,y). \tag{6.28}$$

Similarly, the contribution due to system Y is

$$\left.\frac{\partial}{\partial t'}\right|_{t'=t} \rho_{X(t)Y(t')}(x,y) = \sum_{y'} J^{x}_{y'\to y}(t) \tag{6.29}$$

where

$$J^{x}_{y'\to y}(t) = \Gamma^{x}_{y'\to y}\rho_{X(t)Y(t)}(x,y') - \Gamma^{x}_{y\to y'}\rho_{X(t)Y(t)}(x,y). \tag{6.30}$$

Notice that these two equations, (6.27) and (6.29), contain the terms in the master equation (D.15) corresponding to the jumps on system X and Y, respectively. The sum of the two yields the master equation ruling the evolution of the global probability distribution $\rho(x,y;t) \equiv \rho_{X(t)Y(t)}(x,y)$.

Now, every term in Eq. (6.16) can be decomposed into two parts accounting for the contributions of transitions in systems X and Y, respectively. For transitions in system X, this contribution is

$$\dot{S}^{X}_{\text{prod}}(t) \equiv \dot{S}(X(t)) - \dot{I}^{X}(t) + \dot{S}^{X}_{\text{env}}(t). \tag{6.31}$$

Each derivative here is expressed in terms $J^{y}_{x'\to x}(t)$. Let us discuss the three terms in Eq. (6.31). Using similar arguments as in the derivation of Eq. (D.33), one can obtain expressions for the Shannon entropy

$$\dot{S}(X(t)) = k\sum_{y}\sum_{x>x'} J^{y}_{x'\to x}(t)\ln\frac{\rho_{X(t)}(x')}{\rho_{X(t)}(x)} \tag{6.32}$$

and the information flow

$$\begin{aligned}\dot{I}^{X}(t) &= k\sum_{x,y}\sum_{x'} J^{y}_{x'\to x}(t)\ln\frac{\rho_{X(t)Y(t)}(x,y)}{\rho_{X(t)}(x)\rho_{Y(t)}(y)}\\ &= k\sum_{x,y}\sum_{x'} J^{y}_{x'\to x}(t)\ln\frac{\rho_{X(t)Y(t)}(x,y)}{\rho_{X(t)}(x)}\\ &= k\sum_{y}\sum_{x>x'} J^{y}_{x'\to x}(t)\ln\frac{\rho_{X(t)Y(t)}(x,y)\rho_{X(t)}(x')}{\rho_{X(t)Y(t)}(x',y)\rho_{X(t)}(x)}.\end{aligned} \tag{6.33}$$

The entropy increase in the environment due to jumps in X is given by

$$\dot{S}^X_{\rm env}(t) = k \sum_y \sum_{x>x'} J^y_{x'\to x}(t) \ln \frac{\Gamma^y_{x'\to x}}{\Gamma^y_{x\to x'}}. \qquad (6.34)$$

This expression is a consequence of local detailed balance, as explained in section D.5. Introducing Eqs. (6.32), (6.33), and (6.34) into (6.31), we obtain:

$$\dot{S}^X_{\rm prod}(t) = k \sum_y \sum_{x>x'} J^y_{x'\to x}(t) \ln \frac{\rho_{X(t)Y(t)}(x',y)\Gamma^y_{x'\to x}}{\rho_{X(t)Y(t)}(x,y)\Gamma^y_{x\to x'}} \geq 0. \qquad (6.35)$$

The inequality is valid for every term in the sum since the current and the logarithm have equal signs.

In the stationary state, the two inequalities corresponding to transitions of system X and Y read respectively

$$\dot{S}^X_{\rm prod} = \dot{S}^X_{\rm env} + \dot{I}^Y \geq 0 \qquad (6.36)$$
$$\dot{S}^Y_{\rm prod} = \dot{S}^Y_{\rm env} - \dot{I}^Y \geq 0. \qquad (6.37)$$

If $\dot{I}^Y$ is positive, Y acts as a demon that measures and operates on X. System Y uses the entropic resources of its environment, $\dot{S}^Y_{\rm env} \geq \dot{I}^Y$, whereas system X can reduce the entropy of the corresponding reservoirs: $\dot{S}^X_{\rm env} \geq -\dot{I}^Y$. This use of resources can be represented in a simple flow diagram (also called a Sankey diagram), as the one depicted in figure 6.5. The bipartite is a machine fueled by the Y-environment. The increase of entropy in the Y-environment is transmitted to system X through correlations. The information flow $\dot{I}^Y$ quantifies this transfer and is a resource for system X, which uses it to reduce the entropy of its environment. The losses in each stage are the entropy productions $S^X_{\rm prod}$ and $S^Y_{\rm prod}$. The diagram clearly illustrates that mutual information is equal to the exchange of entropy between two subsystems that comprise a bipartite system.

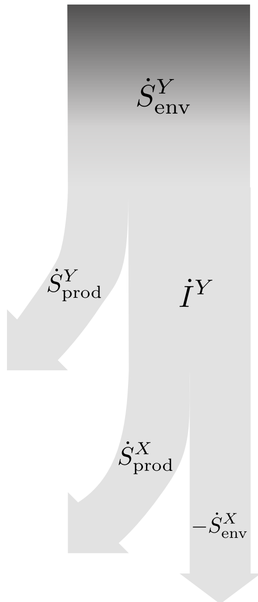


Fig. 6.2 Sankey diagram of a bipartite system.

6.4 A case study: feedback vs. chemical motors

In this section, we illustrate the ideas discussed in this chapter with a simple example of a feedback engine that admits different interpretations depending on the physical mechanism implementing the feedback.

The model consists of a Brownian particle in a potential that alternates between two profiles, A and B, as shown in figure 6.3. The potentials A and B have a staircase shape plus infinite barriers that create separate compartments where the particle moves. The barriers of each potential are shifted with respect to the other. One can induce an uphill motion if the potential is switched from B to A when the particle is in the rightmost step of a compartment generated by potential B (labeled as R in Fig. 6.3) and from A to B when the particle is in the rightmost

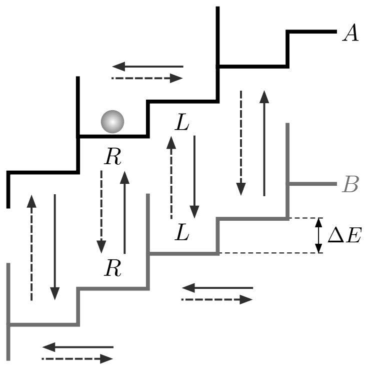


Fig. 6.3 The feedback engine discussed in the text. A Brownian particle moves in a potential that consists of a constant bias and barriers, and switches between two shapes, A and B.

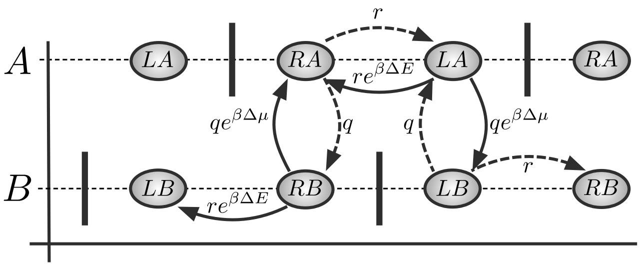


Fig. 6.4 Representation of the feedback engine of Fig. 6.3 as a bipartite Markov chain. Transitions are biased towards the direction indicated by the solid arrow.

step of a compartment generated by potential A (labeled as L in the figure). These switches can be interpreted as transitions between two internal states of the particle, A and B. Fig. 6.3 shows the transitions between these internal states (vertical arrows) and the transitions between compartments induced by the Brownian motion of the particle (horizontal arrows). The figure's solid arrows indicate the direction towards which the transitions are biased. The staircase potential creates a bias towards the left, whose intensity is modulated by the potential energy gain per step, ΔE. The transitions between internal states are biased in a way that favors the uphill motion, as seen in Fig. 6.3, and can be induced by different mechanisms, which we explore in this section.

6.4.1 Information flows in a chemical motor

The first mechanism is a non-equilibrium reaction that mediates transitions between internal states A and B, as explained in section D.4. Notice however that the mediation depends on the position of the particle:

$$RB \overset{C \;\curvearrowright\; C'}{\longleftrightarrow} RA \qquad\qquad LA \overset{C \;\curvearrowright\; C'}{\longleftrightarrow} LB$$

Here C and C' are chemical species from a non-equilibrium chemostat. This means that the concentration of C and C' are fixed to non-equilibirum values yielding a chemical potential difference $\Delta\mu \equiv \mu_C - \mu_{C'} > 0$ that keeps the reactions out of equilibrium and induces a bias towards transitions that convert C into C', providing the fuel that pumps the Brownian particle.

The corresponding rates are depicted in figure 6.4, where we can also check that the resulting Markov chain is bipartite. Since the system is periodic in space, we can focus on a single period described by a bipartite Markov chain with four states, LA, LB, RA, and RB. The rates obey local detail balance, as explained in Appendix D:

$$\begin{array}{ll} \Gamma_{RB\to RA} = qe^{\beta\Delta\mu} & \Gamma_{RA\to RB} = q \\ \Gamma_{RA\to LA} = r & \Gamma_{LA\to RA} = re^{\beta\Delta E} \\ \Gamma_{LA\to LB} = qe^{\beta\Delta\mu} & \Gamma_{LB\to LA} = q \\ \Gamma_{LB\to RB} = r & \Gamma_{RB\to LB} = re^{\beta\Delta E} \end{array} \tag{6.38}$$

where β is the inverse temperature and r and q are rates that determine the time scale of the diffusion and the chemical reaction, respectively. The system is effectively one-dimensional. Then, in the steady state, p_i, the current J is the same at every link:

$$\begin{aligned} J &= p_{RB} q e^{\beta\Delta\mu} - p_{RA} q \\ J &= p_{RA} r - p_{LA} r e^{\beta\Delta E} \\ J &= p_{LA} q e^{\beta\Delta\mu} - p_{LB} q \\ J &= p_{LB} r - p_{RB} r e^{\beta\Delta E}. \end{aligned} \tag{6.39}$$

The solution reads:

$$p_{RB} = p_{LA} = \frac{J(q+r)}{qr(e^{\beta\Delta\mu} - e^{\beta\Delta E})} \tag{6.40}$$

$$p_{RA} = p_{LB} = \frac{J(qe^{\beta\Delta\mu} + re^{\beta\Delta E})}{qr(e^{\beta\Delta\mu} - e^{\beta\Delta E})}. \tag{6.41}$$

The current can be obtained by imposing the normalization of the stationary probability distribution. The result is

$$J = \frac{qr(e^{\beta\Delta\mu} - e^{\beta\Delta E})}{2[q(1+e^{\beta\Delta\mu}) + r(1+e^{\beta\Delta E})]}. \tag{6.42}$$

The sign of the current is the same as the sign of the difference $\Delta\mu - \Delta E$. Hence, the particle moves uphill if $\Delta\mu > \Delta E$, i.e., if the fuel provides sufficient energy to balance the bias due to the potential.

We can now apply the ideas presented in the previous section. Let us denote as $X = L, R$ the position and as $Y = A, B$ the internal state of the particle. These are the two variables of our bipartite Markov chain. They are correlated in the stationary state and their marginal probability distributions are very simple: $p_X(R) = p_X(L) = 1/2$ and $p_Y(A) = p_Y(B) = 1/2$. The entropy production in the stationary regime is the entropy increase in the environment, given by Eq. (D.32), is

$$\begin{aligned} \dot{S}^{(XY)}_{\text{prod}} = \dot{S}_{\text{env}} &= Jk\left[2\ln\frac{qe^{\beta\Delta\mu}}{q} + 2\ln\frac{r}{re^{\beta\Delta E}}\right] \\ &= 2J\,\frac{\Delta\mu - \Delta E}{T} \end{aligned} \tag{6.43}$$

and one can immediately identify $\dot{S}^X_{\text{env}} = -2J\Delta E/T$ as the entropy increase in the thermal bath that generates the Brownian diffusive motion in space, and $\dot{S}^Y_{\text{env}} = 2J\Delta\mu/T$ as the entropy increase in the chemostats due to the changes in the internal state of the particle.

According to (6.33), the information flow reads

$$\begin{aligned} \dot{I}^X = -\dot{I}^Y &= Jk\ln\frac{p_{RA}p_L}{p_{LA}p_R} + Jk\ln\frac{p_{LB}p_R}{p_{RB}/p_L} \\ &= 2Jk\ln\frac{q+r}{qe^{\beta\Delta\mu} + re^{\beta\Delta E}} < 0. \end{aligned} \tag{6.44}$$

We see that $\dot{I}^Y$ is positive, indicating that system Y, the internal state, acts as a demon on system X, the position of the particle. The internal

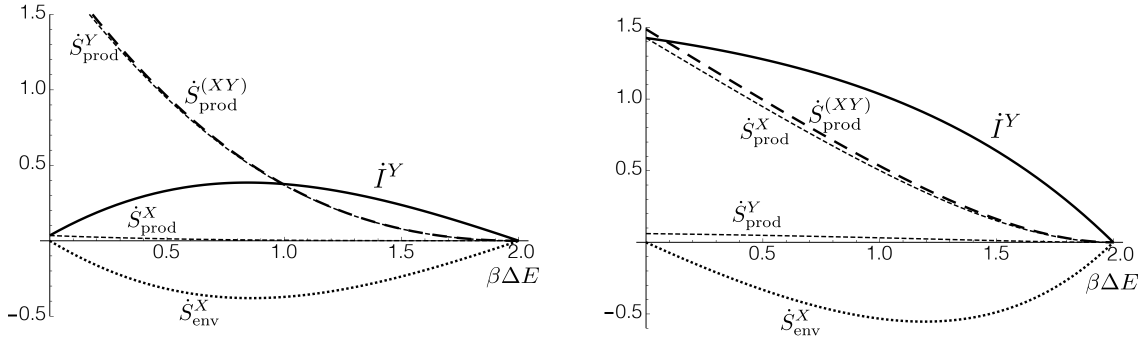


Fig. 6.5 Total entropy production $\dot{S}^{(XY)}_{\rm prod}$ (thick dashed line), information flow $\dot{I}^{Y}$ (solid line), partial entropy productions (dotted lines) for the chemical motor, with $\beta\Delta\mu = 2$ in units of the Boltzmann constant k. The left plot corresponds to the case where diffusion is much faster than the potential switch, $q = 0.3$, $r = 50$, whereas the right plot corresponds to the opposite case, $q = 10$, $r = 1$. We also plot the entropy production due to the diffusion, $\dot{S}^{X}_{\rm env} = -J\Delta E/T$ (lower dotted line), which is negative and proportional to the power of the motor.

state measures the position of the particle, due to the spatial dependence of the reaction. Finally, the entropy production can be split into two terms:

$$\begin{aligned} \dot{S}^{X}_{\rm prod} &= \dot{I}^{Y} - 2J\frac{\Delta E}{T} \\ \dot{S}^{Y}_{\rm prod} &= -\dot{I}^{Y} + 2J\frac{\Delta\mu}{T}. \end{aligned} \tag{6.45}$$

As discussed above, the internal state Y consumes an amount of free energy per unit of time $2J\Delta\mu$ to measure the position of the particle, i.e., to correlate its state with the state of X. These correlations are used to extract an energy ΔE per uphill step from the thermal bath, which is converted into potential energy of the particle. One can assess the efficiency of each process —measurement and correlation exploitation— by comparing the two contributions to the entropy production. For instance, if $q \gg r$, i.e., if the dynamics of the internal state (i.e., the chemical reactions) is much faster than the spatial diffusion, then, according to Eq. (6.44), $\dot{I}^{Y} \simeq 2J\Delta\mu/T$. Hence, $\dot{S}^{Y}_{\rm prod} \simeq 0$ and $\dot{S}^{X}_{\rm prod} \simeq \dot{S}^{(XY)}_{\rm prod} = 2J(\Delta\mu - \Delta E)/T$. In this case, the "measurement" is very efficient and all the entropy is produced by the "feedback" part of the bipartite system. On the contrary, if $q \ll r$, then $\dot{I}^{Y} \simeq 2J\Delta E/T$, and all the dissipation is produced by the measurement. A particular case is shown in figure 6.5, where we depict the different contributions to the entropy production (in units of the Boltzmann constant k) as a function of the load $\beta\Delta E$ for $\beta\Delta\mu = 2$ and different values of the diffusion rate r and the chemical rate q.

6.4.2 Non-autonomous feedback

We now analyze the model as an information engine without any assumption about the physical nature of the demon, in the spirit of our discussion in chapter 4. An external agent measures the position of the particle and chooses potential A or B to induce the uphill motion. This system is similar to the experimental realization of a Maxwell demon reported in (Toyabe, Sagawa, Ueda, Muneyuki and Sano, 2010). To study the system quantitatively, we need to specify the measurement and feedback schedule. It is interesting to consider a particular protocol that induces the same dynamics as the chemical motor discussed above (Horowitz, Sagawa and Parrondo, 2013). In this way, we can compare the chemical motor, interpreted as an exchange of free energy and information between the spatial and the internal states of the particle, with a pure information motor based on feedback.

The dynamics induced by the chemical reaction consists of Poissonian jumps between internal states. The potential is switched at a rate $qe^{\beta\Delta\mu}$ when the particle is in the rightmost half of the container created by the current potential, A or B, and at a rate q when it is in the leftmost half. The latter transition can be interpreted as an erroneous measurement. To reproduce the dynamics of the chemical motor, we assume that the demon measures the position of the particle at random times distributed as Poissonian events with a rate α and an error probability ϵ, such that $\alpha(1-\epsilon) = qe^{\beta\Delta\mu}$ and $\alpha\epsilon = q$, i.e. (Horowitz, Sagawa and Parrondo, 2013):

$$\begin{aligned} \epsilon &= \frac{1}{1+e^{\beta\Delta\mu}} \\ \alpha &= q\left(1+e^{\beta\Delta\mu}\right) \end{aligned} \tag{6.46}$$

Notice that an equilibrium chemostat, $\Delta\mu = 0$, is reproduced by a demon with an error probability $\epsilon = 1/2$ in the measurement.

The demon can take two actions in each small interval of time[2] of duration Δt: either switching the potential (S) or doing nothing (N). The relevant quantity in the second law for feedback processes is the mutual information between the possible actions of the demon, $M = S, N$, and the state of the system. In this analysis, we do not need to specify the internal state A or B, but only the switches. These switches occur when the particle is in the rightmost half of the container formed by the two potential barriers acting at the moment of the measurement. Then, for the state of the system, it is enough to consider two states $X = L, R$, corresponding to the leftmost or rightmost half of the current container, respectively. Notice that this random variable X does not coincide with the one that we have used in the previous subsection to analyze the motor as a bipartite system. The conditional probabilities

[2] There are two ways of generating Poissonian events at a rate α. The first is placing events at a distance $\mathcal{T}$, where $\mathcal{T}$ is a random time whose probability density is exponential: $\rho_{\mathcal{T}}(t) = \alpha e^{-\alpha t}$. The second is discretizing time in small intervals of size Δt and generating an event in each interval with a probability $\alpha\Delta t \ll 1$. Here, we consider the second scheme, which is more suitable for calculating mutual information.

of these two random variables read

$$p_{M|X}(S|R) = \alpha(1-\epsilon)\Delta t \qquad (6.47)$$
$$p_{M|X}(S|L) = \alpha\epsilon\Delta t$$

and $p_{M|X}(N|R) = 1 - p_{M|X}(S|R)$, $p_{M|X}(N|L) = 1 - p_{M|X}(S|L)$. Then, the mutual information between the demon and the system after each measurement attempt in an interval of duration Δt is[3]

$$\begin{aligned} I(X;M) &= S(M) - S(M|X) \\ &= S(p_s\alpha\Delta t) - p_X(R)S(\alpha(1-\epsilon)\Delta t) - p_X(L)S(\alpha\epsilon\Delta t) \end{aligned} \qquad (6.48)$$

where $S(p)$ is the Shannon entropy of a binary variable, as defined in (2.2), and p_s is the probability of switching in a given measurement:

$$p_{\mathrm{s}} \equiv \frac{p_M(S)}{\alpha\Delta t} = p_X(R)(1-\epsilon) + p_X(L)\epsilon. \qquad (6.49)$$

Since, for $p \ll 1$, $S(p) \simeq p - p\ln p$, Eq. (6.48) for $\Delta t \to 0$ can be simplified to

$$I(X;M) \simeq \alpha k\Delta t \left[p_X(R)(1-\epsilon)\ln\frac{1-\epsilon}{p_s} + p_X(L)\epsilon\ln\frac{\epsilon}{p_s} \right]. \qquad (6.50)$$

To calculate the probability distribution p_X of the position of the particle we can use the results of the chemical motor since the dynamics of both models is the same. Hence

$$p_X(R) = \frac{q+r}{q(1+e^{\beta\Delta\mu}) + r(1+e^{\beta\Delta E})} \qquad (6.51)$$

and $p_X(L) = 1 - p_X(R)$. The entropy production rate finally reads

$$\dot{S}^{\mathrm{fb}}_{\mathrm{prod}} = \lim_{\Delta t\to 0}\left[\frac{I(X;M)}{\Delta t}\right] - \frac{2J\Delta E}{T}. \qquad (6.52)$$

[3] One could also calculate the mutual information between the state of the system X and the outcome of the measurement $\mathcal{M} = L, R$: $I(X;\mathcal{M}) = S(\mathcal{M}) - S(\mathcal{M}|X) = h(p_s) - h(\epsilon)$. Then the average information gain in an interval of duration Δt is $\alpha\Delta t\, I(X;\mathcal{M})$, which is larger than the mutual information calculated in Eq. (6.50). To understand the differences between these two different ways of calculating the mutual information between the system and the demon, consider the following scenario: a measurement is carried out in every interval of time Δt, with three possible outcomes $\mathcal{M}' = \emptyset, L$, and R: we obtain $\emptyset$ (null measurement) with probability $1-\alpha\Delta t$ independent of the state of the system, and R and L with probabilities $p_{\mathcal{M}'|X}(m|x) = \alpha\Delta t\,(1-\epsilon)$ if $x \neq m$ and $p_{\mathcal{M}'|X}(m|x) = \alpha\Delta t\,\epsilon$ if $x = m$. The mutual information is $I(X;\mathcal{M}') = S(\mathcal{M}') - S(\mathcal{M}'|X) = \alpha\Delta t I(X;\mathcal{M}) = \alpha\Delta t(h(p_s) - h_\epsilon)$. On the other hand, Eq. (6.50) yields the value of the mutual information between the system and the *actions* that the demon can take, which are only two: N and S. These two actions are the result of lumping the measurement outcomes $\mathcal{M}' = \emptyset, R$ into a single action N, i.e.: $M = N$ if $\mathcal{M}' = \emptyset, R$, and $M = S$ if $\mathcal{M}' = L$. Then, it is clear that $I(X;M) \leq I(X;\mathcal{M}')$ (see Exercise 6.4 for a rigorous proof). To assess the performance of the motor, we use the actions M and the mutual information (6.48), since they comprise all information used in the feedback.

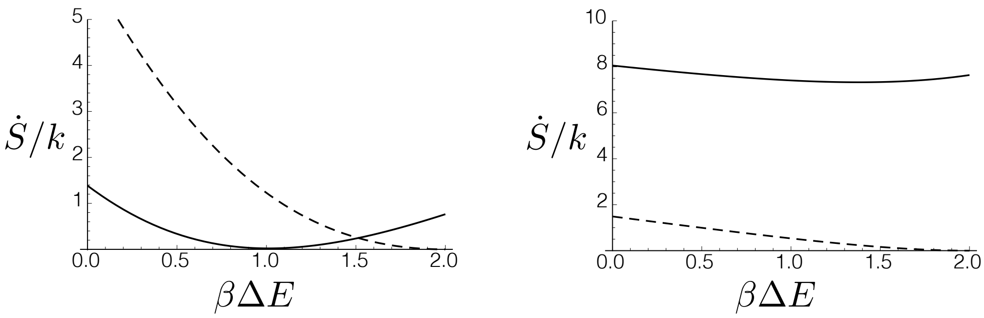


Fig. 6.6 Entropy production $\dot{S}^{\rm fb}_{\rm prod}$ and $\dot{S}^{(XY)}_{\rm env}$, for the feedback and the chemical motor respectively, with $\beta\Delta\mu = 2$. The leftmost plot corresponds to the case where diffusion is much faster than the potential switch, $q = 1$, $r = 100$, whereas the rightmost plot corresponds to the opposite case, $q = 10$, $r = 1$.

We can compare this entropy production with the entropy production of the autonomous system, given in (6.43): $\dot{S}^{(XY)}_{\rm prod} = 2J(\Delta\mu - \Delta E)/T$. We show two interesting cases in Fig. 6.6. The most remarkable feature can be seen in the leftmost plot, where the entropy production in the feedback engine is zero. This occurs for $\beta\Delta E = 1$ and $\beta\Delta\mu = 2$ and $r \gg q$. For these values the power of the machine is finite:

$$\mathcal{P} = 2J\Delta E = \frac{qr\Delta E(e^2 - e)}{q(1+e^2) + r(1+e)} \simeq q\Delta E \tag{6.53}$$

, i.e., the machine can transport particles against the bias ΔE at finite power with zero entropy production.

Exercises

(6.1) The mild Szilárd paper Mass

(6.2) *Data-processing inequality (Cover and Thomas, 2006).* Consider three random variables $X, \mathcal{M}'$, and M that form a Markov chain $X \leftrightarrow \mathcal{M}' \leftrightarrow M$, i.e., $\rho_{M|\mathcal{M}'X}(m|m', x) = \rho_{M|\mathcal{M}'}(m|m')$. Prove that $I(X;M) \leq I(X;\mathcal{M}')$ and that the equality is reached if $M = f(Y)$. Apply this to the system discussed in section 6.4.2, where $X = L, R$ is the position of the particle, $\mathcal{M}' = \emptyset, L, R$ is the outcome of the measurement and $M = N, S$ is the action taken by the demon (see footnote 3 in page 108 and figure 6.7).

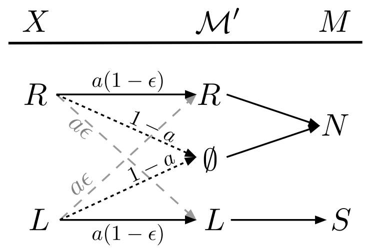


Fig. 6.7 Random variables involved in the feedback motor: the position of the particle X, the outcome of the measurement $\mathcal{M}'$ (where $\emptyset$ is the absence of measurement), and the action M taken by the demon. Here $a = \alpha\Delta t$ is the probability of measuring in an interval of duration Δt and ϵ is the error probability (see footnote 3). The conditional probability $p_{\mathcal{M}'|X}(m'|x)$ is given by the arrows connecting the values of X and $\mathcal{M}'$: solid arrows indicate correct measurements, gray dashed lines indicate wrong measurements, and dotted line indicate null measurements.

(6.3) Prove that Eq. (6.22) implies Eq. (6.23). Hint: calculate the conditional average

$$\left.\frac{d}{dt'}\right|_{t'=t} \left\langle A(X(t'))B(Y(t'))\,|\,x, y, ; t\right\rangle .$$

where $A(x)$ and $B(y)$ are arbitrary functions.

(6.4) Consider two stochastic processes $X(t)$ and $Y(t)$. Prove that the derivative of the mutual information between $X(t)$ and $Y(t')$ can be written as

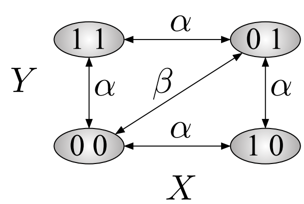
$$\begin{aligned}\frac{d}{dt}I(X(t);Y(t')) &= k\int dxdy\,\frac{\partial\rho_{X(t)Y(t')}(x,y)}{\partial t} \\ &\times\ \ln\frac{\rho_{X(t)Y(t')}(x,y)}{\rho_{X(t)}(x)\rho_{Y(t')}(y)}\end{aligned}$$

Use this result to prove Eq. (6.19) from Eq. (6.23).

(6.5) Prove that a bipartite Markov chain verifies Eq. (6.23).

(6.6) Prove Eq. (6.27) for a bipartite system.

(6.7) Consider the following Markov chain that describes the state of two binary systems $X, Y = 0$ or 1:



α and β being transition rates. Prove that Eq. (6.23) is satisfied if and only if $\beta = 0$.