

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Fundamentals on Energy in ICT

---

Luca Gammaitoni

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/66973>

---

## Abstract

This chapter deals with the fundamental physical aspects of the use of energy in ICT devices. Here we discuss questions such as “what is the theoretical minimum energy required to process information?”, “what is the minimum energy required to transmit information from one point to another?” and “are these limits practically reachable and under what conditions?” While dealing with these relevant questions, we are mostly concerned with providing to the reader a clear and intuitive understanding of what is going on and what the underlying physics aspects are.

**Keywords:** energy, information, noise, efficiency, entropy

---

## 1. Introduction

This chapter deals with the fundamental physical aspects of the use of energy in ICT devices. Here, we discuss questions like *what is the theoretical minimum energy required to process information? What is the minimum energy required to transmit information from one point to another? And are these limits practically reachable and under what conditions?*

Most importantly, in dealing with these relevant questions, we will be mostly concerned with providing to the reader a clear and intuitive understanding of what is going on and what are the underlying physical aspects, more than showing rigorous mathematical demonstrations. In fact, these can be found in many university textbooks (some listed at the end of the chapter) and missing some rigor will hopefully not harm the validity of the reasoning.

Dealing with fundamentals in ICT necessarily implies dealing with physics. In fact any ICT device, being it a complex microprocessor with billions of transistors interconnected or a simple binary logic gate is, first of all, a physical system. As such its functioning is subjected to the laws of physics. Regarding the implications of the use of energy in such devices we are

---

thus referred to the very elegant theory of thermodynamics. In this theory, many scientists through the years have accumulated all the knowledge developed in dealing with energy and its transformations. Thanks to the work of scientists like Sadi Carnot, Emile Clapeyron, Rudolf Clausius and William Thomson (Lord Kelvin), studies on how energy could be used with profit in machines invented earlier by Thomas Newcomen and James Watt to transform heat into work, brought us the notion of entropy and the second law of thermodynamics that put limits on the efficiency of such machines.

Steam engines from the dawn of the industrial revolution are not much different from nowadays ICT systems if you look at them by a merely physics point of view. In both classes of devices, we are dealing with the transformation of energy from heat to work and from work to heat. Surprisingly, 200 years have passed after the work of Carnot but we are still intrigued by the problem of defining the efficiency of these transformations even if, at difference with the past, today the object of our quest has moved from the heat engines of the industrial revolution to the tiny devices of modern ICT.

It is a common statement that future ICT will be characterized by nanoscale devices that will process information while dissipating significant amount of energy, i.e. while transforming work into heat. In this perspective it seems natural to consider an ICT device as a novel info-thermal machine: it inputs information and energy (in the form of work) and outputs information and energy (in the form of heat).

In the following, we will discuss in detail these aspects trying to make clear what are the underlying fundamental physical laws that govern the use of energy in ICT devices. We will proceed as follows:

2. What is information processing and how this can be done with machines?
3. Basics on thermodynamics laws.
4. Digital computing and the physics of switches.
5. Energy efficiency, Landauer reset and reversible computing.
6. Energy bounds on communication as information transfer processes.

## **2. What is information processing and how this can be done with machines?**

In this section, we discuss the fundamentals in information processing. We introduce the notion of ‘amount of information’ and its digital representation. Most importantly, we discuss how a physical system can be used to do information processing and how this has to do with the laws of physics and with energy transformation processes in particular.

Let us start with a fundamental question: what does it mean ‘information processing’? Before we can answer this question, we need to introduce the notion of ‘information’.

This notion was introduced, in the framework that is of interest here, by Claude Shannon (1916–2001) in 1948 in his attempt to formulate a ‘mathematical theory of communication’.

During a communication process, there is a message that needs to be transmitted from one point to another. In this perspective, the ‘amount of information’ is a quantity that can be associated with a given message.

To illustrate this concept, let us assume that we want to transmit a text message. Something like ‘Hello my friend, what’s up?’. This message is composed of a number letters and punctuation symbols. Let us suppose that this message is part of a much longer message so that we can assign to each symbol a given probability to be part of this message. Typically, if the message is written in a given language, we can use the frequency of each letter in that language as the probability. If we call a generic symbol  $x_i$  (this can be a letter like ‘a’ or ‘A’ or a punctuation symbol like ‘;’) then we can indicate the probability to find it in our message is  $p(x_i)$ . At this point, we can define the amount of information that is carried by each symbol  $x_i$  (according to Claude Shannon definition) as the number  $H_i$  given by:

$$H_i = -K p(x_i) \log p(x_i) \quad (1)$$

where  $K$  is simply a constant and ‘log’ represents the logarithmic function. By the moment that the probability  $p(x_i)$  is a number between 0 and 1, the  $\log p(x_i)$  is a negative number and thus the resulting  $H_i$  is a positive quantity.  $H_i$  is also sometimes called ‘entropy’, in analogy with the physics quantity as introduced by Gibbs previously (see below).

When we want to transmit a message, it is more practical to transmit only a small number of different symbols in order to avoid possible confusion between two similar symbols. In fact, you can easily realize that it is easier to transmit the Latin alphabet, with say 25 symbols, than the Japanese katakana with 48 different characters. Thinking about it, people realized that the most confusion-avoiding (noise-resistant) way to do this is to associate each symbol (character) you want to transmit to a number and then to represent the number in base 2, with only two different digits, e.g. ‘0’ and ‘1’.

In synthesis, when we transmit a message we transmit a stream of symbols ‘0’ and ‘1’ that can be associated with numbers that are associated with letters and punctuations. Traditionally, these symbols ‘0’ and ‘1’ are called *bit* as a contraction of the words ‘binary digit’.

In most common long messages, the probability to find a ‘0’ or a ‘1’ is the same and is thus  $p(0) = p(1) = 0.5$ . According to this, the amount of information transported by a message composed of  $n$  bits with  $n_0$  symbols ‘0’ and  $n_1$  symbols ‘1’ (and  $n = n_0 + n_1$ ) is:

$$H = -K \left( n_0 \frac{1}{2} \log \frac{1}{2} + n_1 \frac{1}{2} \log \frac{1}{2} \right) = K \frac{1}{2} (n_0 + n_1) \log 2 \quad (2)$$

If we chose  $K = 2$  and assume that  $\log 2 = 1$  (this is true if we admit the base 2 for the logarithmic function), we have

$$H = (n_0 + n_1) = n \quad (3)$$

Thus, the amount of information in a message coded in bits is equal to the number of bits in the message.

Within this framework, ‘information processing’ is what we do when we manipulate (= do operations on) the bits of a message. By the moment, bits are numbers, ‘information processing’ is substantially equivalent to computing.

So we are back to another fundamental question: how do we do computing? This question may seem a bit naïve by the moment that we are all used to deal with the act of computing since when we were kids. Clearly computing is associated with dealing with quantities, represented by numbers: how to count them, how to add, subtract and in general how to transform numbers. This is absolutely correct. However, here we would like to focus on the fact that, at its very fundamental roots, the act of computing can be associated with very simple physical manipulations like moving a ball from one pot to another or changing the position of a pebble in a row or a column. This was known since the old times: the word ‘calculus’ (form which ‘calculations’) comes from Latin and designated small stones that Romans used to account for quantities, i.e. to perform computations.

Manipulating physical objects is thus at the base of computing and we have shown that such manipulation has the power of transforming not only numbers but more generally any kind of symbol, as it is normally carried out in modern ICT devices that process information, like when we read an e-mail or change an image on a screen.

By the moment that information processing/computing can be associated with the change of bits, in order to perform this activity we need two very important components:

1. a physical system capable of assuming two different physical states
2. a way to induce state changes in this physical system (typically a force).

We are not going to spend time dealing with the quite philosophical definition of physical system by intending with it any object, device or phenomena that can be studied by physics. The notion of physical state is slightly more delicate. With it we mean a set of measurable quantities whose value can be used to distinguish unambiguously two different outcomes, as an example shown in **Figure 1**.

Here we have

1. The physical system, made by a pebble and two bowls. The two states are represented by the measurable quantity ‘position of the pebble’: state ‘0’ = pebble in left bowl; state ‘1’ = pebble in the right bowl;
2. The way to induce state changes represented by a force that brings around the pebble.

According to this example, we can perform information processing activity simply by changing the position of the pebble, according to certain rules, with the underlining idea that, while we do these changes, we are at the same time changing the value of the symbol ‘0’ and ‘1’ associated with the system state. Devices that obey the rules (1) and (2) are called *binary switches*.

In modern computers, binary switches are made of transistors. These are electronic devices (**Figure 2**, left) that satisfy the two required conditions:

1. The two states are represented by the measurable quantity 'electric voltage' at point  $V_{OUT}$ . As an example state, '0' =  $V_{OUT} < V_T$ ; state '1' =  $V_{OUT} > V_T$ ; with  $V_T$  a given reference voltage;
2. The way to induce state changes represented by an electromotive force applied at point  $V_{IN}$ .



Figure 1. Two bowls and one pebble are sufficient to realize a binary switch.

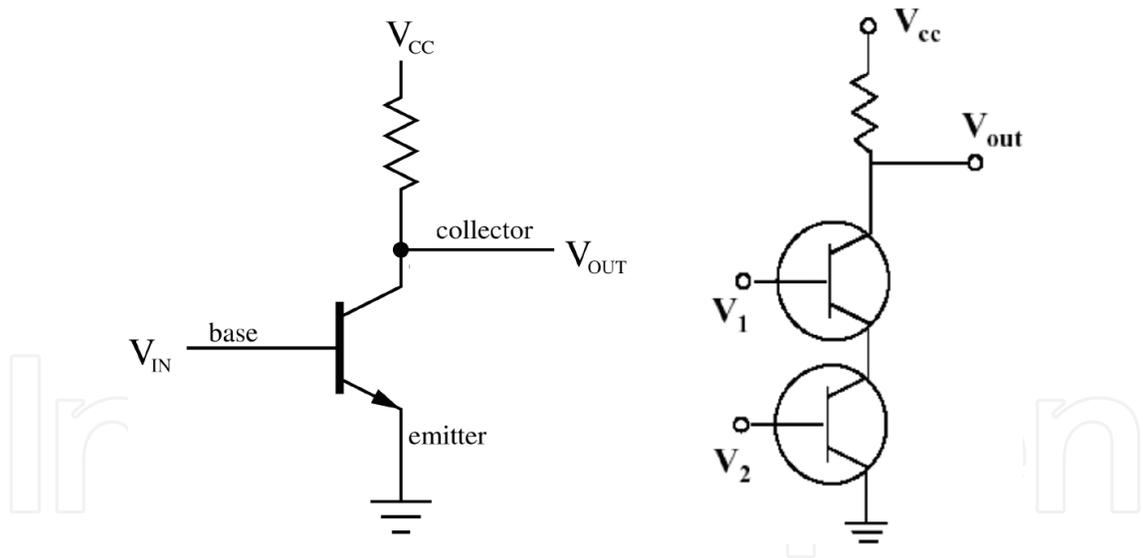


Figure 2. Left: transistor. Right: a combination of two transistors to realize a NAND gate.

By combining binary switches, we can perform all the information processing operations required. As an example, we mention the NAND gate (a universal logic gate) that can be realized by interconnecting two transistors (see **Figure 2**, right).

Now, the question that we want to address is the following: what is the minimum energy required to process information? In order to answer this question, we have to briefly recall the basic laws of one of the most elegant physics theories: thermodynamics.

### 3. Basics on thermodynamics laws

Thermodynamics is the theory that deals with concepts like energy, work, heat, entropy and their use in physical systems. In this section, we present in a concise way the fundamental laws of thermodynamics [1]. It will help us to understand what can we do and what we cannot do with energy.

The fundamental laws were considered through a period of approximately 100 years during which wrong assumptions, brilliant experiments and hard work characterized the work of a bunch of great scientists. Among them we list Thomas Newcomen (1664–1729) who built the first practical steam engine aimed at pumping water out of coal mines and James Watt (1736–1819) who soon after realized an improved version of the same machine. The laws of thermodynamics were considered to provide understanding and tools to the engine makers. This effort was carried out in few decades by some remarkable scientist: Émile Clapeyron (1799–1864), Sadi Carnot (1796–1832), Rudolf Clausius (1822–1888), and William Thomson (Lord Kelvin) (1824–1907).

The laws of thermodynamics do not tell us much about what energy is but they are very good in ruling what can we do and what we cannot do when we change the energy content of a body by exchanging heat and work.

*The first law* of thermodynamics is about the *conservation of energy*. It states that the total energy of a physical system remains the same during any transformation the system can go through, provided we take into account how much work the system does and how much heat the system exchanges (i.e. work and heat balances out).

It was first proposed by Julius Robert von Mayer (1814–1878) and subsequently reviewed by James Prescott Joule (1818–1889) and Hermann Ludwig Ferdinand von Helmholtz (1821–1894). Conservation of energy is strongly believed to be true and, to some extent, it is a self-sustaining law: it is so strongly believed that in every instance we observe a possible violation, we think harder to discover some way in which energy could have been overlooked and if we cannot find a way, well... we invent a new kind of energy. In past we did so at the beginning of 1900 when Albert Einstein introduced the mass-energy equivalence, to account for the 'missing mass' during a nuclear transformation.

*The second law* is about how much energy in the form of heat we can draw from a system in order to do work. Specifically, the second law shows that there are limitations to the amount of work we can get from a given amount of energy present in the form of heat. There are few equivalent formulations of these laws. We list here the two most popular, ascribed to Rudolf Clausius and Lord Kelvin:

Clausius: 'No process is possible whose sole result is the transfer of heat from a body of lower temperature to a body of higher temperature'.

Kelvin: 'No process is possible in which the sole result is the absorption of heat from a reservoir and its complete conversion into work'.

An important consequence of the second law, discovered by Sadi Carnot in 1824 (when he was 28 years old), is that there is a limit to the efficiency of a thermal machine. In the publication

entitled *Réflexions sur la Puissance Motrice du Feu* ('Reflections on the Motive Power of Fire') Carnot generalized the concept, popular at that time, of 'steam engine' by introducing the novel concept of 'thermal machine'. A thermal machine is a physical system that can exchange heat and work with its surroundings. Carnot showed that the efficiency of any thermal machine operating between two temperatures is bounded by a quantity that is a function of the two temperatures only and does not depend on the features of the machine (nor the material, nor the geometry, nor the functioning principles). It was a great result, indeed.

Soon after the work of Sadi Carnot, Rudolf Clausius used his result to introduce a new physical quantity that is useful in describing exactly how much heat can be changed into work during the transformation. He suggested the name 'entropy' for this quantity.

The reasoning behind the introduction of this quantity, the entropy, is the following: to operate a thermal machine, it is necessary to find a cyclic transformation during which heat is changed into work. The cycle is necessary because you want to operate the machine continuously and not just once: every time that a cycle is completed you get some work. By reiterating the cycle you can get any amount of work you need. First of all, Clausius proved a theorem that states that during a cyclic transformation, if you do the transformation carefully enough not to lose any energy in other ways (like friction), then the algebraic sum of the heat exchanged with the external (considered positive the heat that goes into the system and negative the heat that leaves the system) divided by the temperature at which the exchanges occur is zero:

$$\oint \frac{dQ}{T} = 0 \quad (4)$$

An important aspect is that the cycle does not depend on the specific path that you take. Moreover, being a cycle, you start and end at the same state. Clausius concluded that, from the previous integral, it does exist a state function  $S$  defined as

$$S_B - S_A = \int_A^B \frac{dQ}{T} \quad (5)$$

(or in differential form  $dS = dQ/T$ ). The function  $S$  is a state function (it only depends on the state) and represents a novel physical quantity called entropy.

In addition, Clausius showed that if during the transformation, you are not careful enough and you lose energy (due to friction), then the inequality holds instead of the equality:

$$\oint \frac{dQ}{T} \leq 0 \quad (6)$$

The transformation like this is also called an *irreversible transformation*. It is easy to show that if we take an irreversible transformation to compute the entropy, we end up with underestimating the change:

$$S_B - S_A \geq \int_{A \text{ irr}}^B \frac{dQ}{T} \quad (7)$$

It is important to point out that in practical cases it is practically unavoidable to have some kind of friction, thus, the inequality holds. In the very special case, in which we have

transformation where we do not have any heat exchange (sometimes called adiabatic transformation), then the right hand of the inequality is zero and the final entropy is always larger than the initial one.

The concept of irreversible transformation is a bit tricky. You can have an irreversible transformation even if there is no apparent friction. This is the case, for example, of the so-called free expansion of a gas. James Prescott Joule in 1845 has shown with a remarkable experiment that you can have a gas to expand freely (without doing any work) from a smaller container to a larger one, without any heat exchange. In this case, the irreversibility of the transformation comes from the fact that during the free expansion, the gas is out of equilibrium, i.e. the usual thermodynamic quantities like temperature, pressure and volume are not well defined due to the fact that the gas is expanding and while parts of the gas are still at a certain temperature (with a given mean velocity), other parts of the gas may show different mean velocities.

If we consider an infinitesimal transformation we have:

$$dS \geq \frac{dQ}{T} \text{ or } TdS \geq dQ \quad (8)$$

where the equal sign holds during a reversible transformation only. The previous equation is often considered the formulation of the second law of thermodynamics [1]. By putting in contact, a physical system that is at temperature  $T_1$  with a heat reservoir that is at temperature  $T_2 > T_1$ , then some heat will be transferred from the reservoir to the system. Accordingly, the integral is positive and the entropy of the system increases (meaning that this process can occur without any work). The other way around phenomenon by which heat is transferred from the system to the reservoir does not occur because it would require a decrease of entropy (second principle) and thus we conclude that during a spontaneous transformation (i.e. without external work) the entropy always increases. We can make the entropy of our system decrease (e.g. like in a refrigerator) but we have to add work from outside [1].

Another way of looking at this formulation of the second principle is the following. In the general case of irreversible transformation, instead of using the inequality, we can write the previous expression as:

$$TdS = dQ + E_d \quad (9)$$

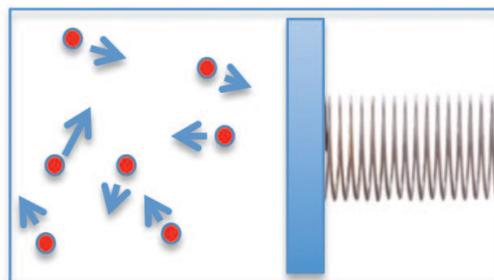
where  $E_d$  is the additional energy dissipated during the transformation, meaning that, when we want the entropy to decrease a quantity  $TdS$ , we need to spend an amount of minimum energy equal to  $dQ$ . If we cannot do things carefully enough to reach a reversible (i.e. lossless, where the quantity  $E_d$  is always zero) transformation condition, then we need to spend  $dQ + E_d$ . Instead, if we want to do the transformation where the entropy increases, then we do not need to spend any minimum amount of energy. Entropy increase can come for free!

Back to the Clausius inequality, it is useful to interpret the quantity  $TdS$  in a reversible transformation as the amount of heat (meaning thermal energy) that cannot be used to produce work [1]. In other words, during the transformation, even if we are carefully enough not to waste energy in other ways, we cannot use all the energy that we have to do useful work, part of this energy

will go into the entropy change. If we are not careful enough the situation is even worse and we get even less work. The limitation in heat transformation is usually quantified by the introduction of the so-called *free energy*. The concept of free energy was proposed by Helmholtz in the form:  $F = U - TS$ . The free energy  $F$  quantifies the maximum amount of energy that we can use to do useful work, when we have available the internal energy  $U$  of a system with entropy  $S$ .

The introduction of entropy was aimed at quantifying the limitations on the use of heat to produce work. However, it is not an exaggeration to say that entropy, in general, remained for many years an obscure quantity, whose physical sense was difficult to grasp. It was the work of Ludwig Boltzmann (1844–1906) that shed some light on the microscopic interpretation of the second law (and thus the entropy). Boltzmann proposed an interpretation of the second law, i.e. the natural tendency of systems to evolve (via spontaneous transformations) towards the state characterized by the increased entropy as the tendency of a system to attain an equilibrium condition identified as the most probable state, among all the states the system can be in.

In the idealized world considered by Boltzmann, physical systems are gasses made by many small parts represented by colliding little spheres. Let us consider an ideal gas made by  $N$  such particles in the form of tiny hard spheres of mass  $m$  that can collide elastically (thus conserving kinetic energy and momentum). Let us suppose that these particles are contained in a box with one of the walls consisting in a moving set of mass  $M = Nm$ . The set is connected to a spring of elastic constant  $k$ , as shown in **Figure 3**, and is at rest [1].



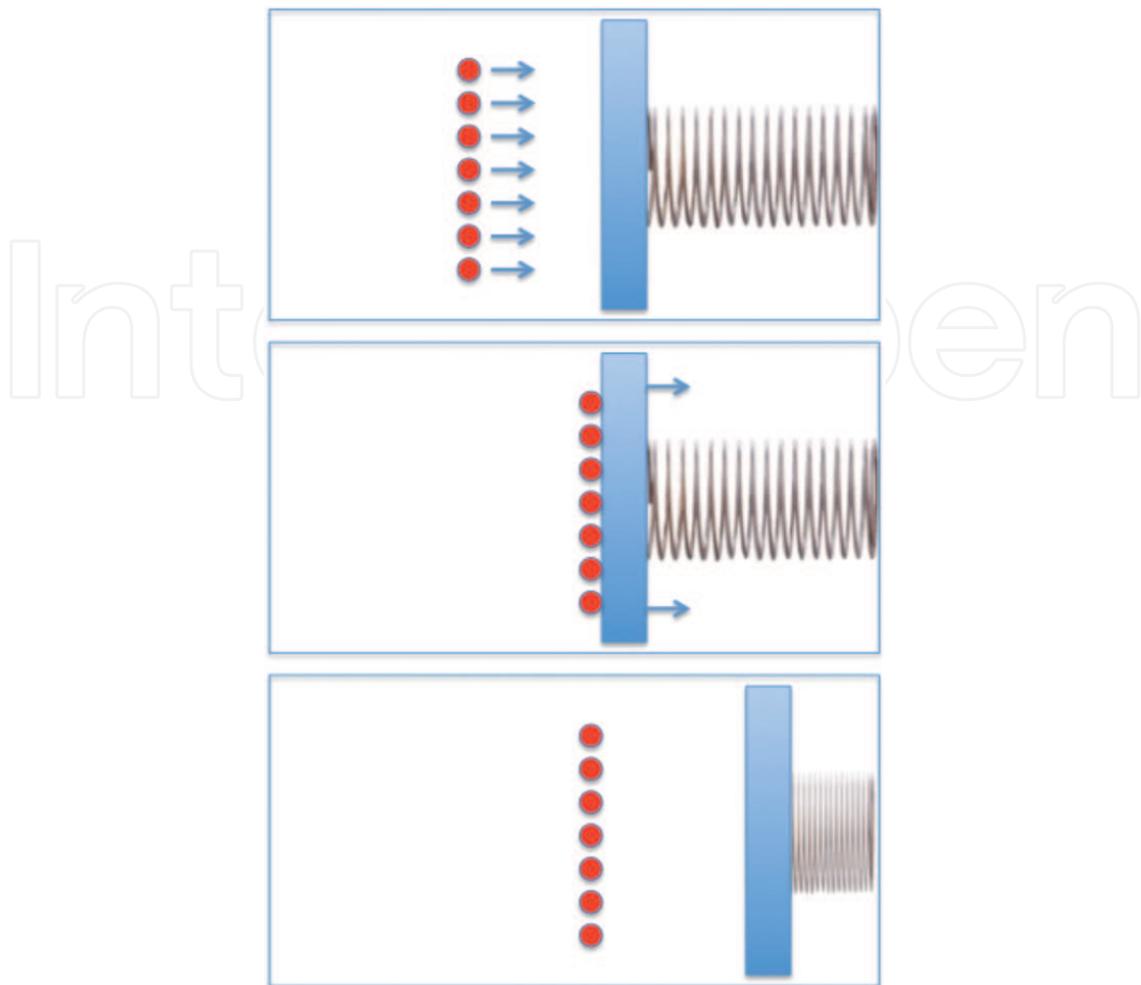
**Figure 3.** Pictorial representation of an ideal gas contained in a box with a moving set connected to a spring. The gas particles move randomly.

If all the particles have the same velocity  $v$  and collide perpendicularly with the moving set at the same time (see **Figure 4**), they will exchange velocity with the set.

This will compress the spring up to an extent  $x_1$  such that:

$$\frac{1}{2}M v^2 = \frac{1}{2}k x_1^2 = U \quad (10)$$

By a purely mechanical point of view, this is a mere transformation of kinetic energy into potential energy. We can always recover the potential energy  $U$  when we desire and use it to perform work. The work will be exactly  $U$ . In this case, we can completely transform the kinetic energy of the gas particle into work. How comes? Well, in this case we are considering



**Figure 4.** Pictorial representation of the ideal gas. Top: all the particles move in ordered way from left to right with parallel velocities. Middle: the particles hit the piston and exchange velocity with it. Bottom: The piston compresses the spring and reach zero velocity. All the energy is now stored as potential energy.

a very special configuration of our gas (unique indeed) where all the particles are moving accordingly in parallel lines. If we put randomly the particles in the box, what is on the contrary the most probable configuration for their arrangement? Based on our experience (and on some common sense as well), the most probable configuration is one where all the particles are moving with random direction (but same velocity) in the box. The kinetic energy of the gas is still the same (so is its temperature  $T$ ) but in this case, the movable set will be subjected at random motion with an average compression of the spring such that its average energy is  $U/N$ . This is also the maximum work that we can recover from the potential energy of the movable set. Thus, it appears clear that, although the total energy  $U$  is the same in the two cases, in the second case we have no hope of using the greatest part of this energy to perform useful work. As we have determined, when we introduced the definition of free energy, the quantity that limits our capability of performing work is the entropy. Following this definition, the system that has the smaller entropy has the larger capability of performing work. Accordingly, we can use the entropy to put a label on the useful energetic content of a system. Two systems may have the same energy but the system that has the lower entropy will have the 'most useful' energy.

This example helped us to understand how energy and entropy are connected to the microscopic properties of the physical systems. In the simple case of an ideal gas, the system energy is nothing else than the sum of all the kinetic energies of the single particles. We can say that the energy is associated with 'how much' the particles move. On the other hand, we have seen that there is also a 'quality' of the motion of the particles that is relevant for the entropy. We can say that the entropy is associated with 'the way' the particles move. This concept of 'way of moving' was made clear by Boltzmann at the end of 1800, who proposed for the entropy the following definition:

$$S = k_B \log W \quad (11)$$

where  $k_B$  is the famous Boltzmann constant and  $W$  is called the 'number of configurations' and represents the number of ways we can arrange all the particles in the system without changing its macroscopic properties. In the previous example, we have only one way to arrange the  $N$  particles so that they are all parallel, aligned and with the same velocity while we have a very large number of ways of arranging the  $N$  particles to be a randomly oriented set of particles with velocity  $v$ . Thus, it is clear that in the second case, the value of the entropy is much larger than that in the first case (where it is indeed zero).

The Boltzmann formula refers to a case where all the microstates are equiprobable. The extension to the more general case with microstates with different probabilities was proposed by Josiah Willard Gibbs (1839–1903):

$$S = -k_B \sum_i p_i \log p_i \quad (12)$$

where  $p_i$  represents the probability of the microstate  $i$ .

We have seen above that during a spontaneous transformation, the entropy of the system increases. This can occur without any change in the energy of the system itself as it was shown by Joule in the famous experiment of free gas expansion. Let us consider our previous example where all the particles move along parallel lines. Let us suppose that the trajectories are not perfectly aligned. Initially nothing happens but, due to a small misalignment, sooner or later a collision between the particles can happen and, after collision, the entire group of particles evolves into a randomly moving group. This is clearly a spontaneous transformation. By the moment that the collisions are elastic the energy of the system has not changed but the system entropy has rapidly increased from the zero initial value up to its maximum value. Conversely, the free energy has reached its minimum value. It is interesting to ask: can we bring the system back to its initial condition? The answer is yes but in order to do it we need to spend some energy as required by the second principle. How much? Clearly we need to spend a minimum of  $T\Delta S$  of energy, where  $\Delta S$  represents the difference in entropy between the final and the initial states. The bad news is that if we spend this energy and decrease the entropy back to its original condition, the energy that we spend does not change the total kinetic energy of the system that remains the same. However, having reduced the system entropy we have increased the Free energy and this improves our capability of extracting work from the system itself.

## 4. Digital computing and the physics of switches

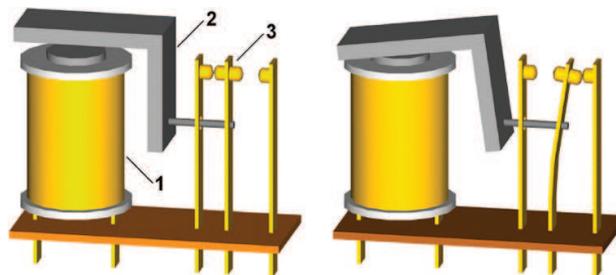
In order to answer our initial question (*what is the minimum energy required to process information?*) we need to apply the thermodynamics concepts that we have just learned to the binary switches that are the basic elements of any information processing device.

As we have discussed in Section 2, a binary switch is a physical system that obeys the two rules (1) and (2) that here we restate as follows:

1. a physical system capable of assuming two different physical states:  $S_0$  and  $S_1$
2. a set of forces that induce state changes in this physical system:  $F_{01}$  produces the change  $S_0 \rightarrow S_1$  and  $F_{10}$  produces the change  $S_1 \rightarrow S_0$ .

If we think about it, we can easily realize that there exist at least two classes of devices that can satisfy these rules. We call them *combinational* and *sequential* devices [1].

*Combinational* devices are characterized by the following behaviour: when no external force is present, under equilibrium conditions, they are in the state  $S_0$ . When an external force  $F_{01}$  is present, they switch to the state  $S_1$  and remain in that state as long as the force is present. Once the force is removed they come back to the state  $S_0$ . Popular examples are represented by relays (**Figure 5**) and also by transistors, today widely exploited in modern computing devices to make logic gates. A combinational device is a network of combinational switches.



**Figure 5.** Electro-mechanical switch (relay) as a combinational device: when an external force (magnetic induction force) is applied the switch changes its state from open to close) and goes back to the initial state (open) when the force is removed. Picture obtained from Wikipedia: [https://commons.wikimedia.org/wiki/File:Relay\\_principle\\_horizontal.jpg](https://commons.wikimedia.org/wiki/File:Relay_principle_horizontal.jpg).

*Sequential* devices are characterized by the following behaviour: if they are in the state  $S_0$ , they can be changed into the state  $S_1$  by applying an external force  $F_{01}$ . Once they are in the state  $S_1$  they remain in this state when the force is removed. The transition from state  $S_1$  to  $S_0$  is obtained by applying a new force  $F_{10}$ . In contrast to the combinational device, the sequential device remembers its state after the removal of the force. This memory lasts for a time that is short compared to the system relaxation time. In fact, if one waits long enough, the sequential device relaxes to equilibrium that, in a symmetric binary switch, is characterized by a 50% probability to be in the  $S_0$  state and 50% probability to be in the  $S_1$  state. This relaxation process is unavoidable in any real physical system that is operated at finite temperature. However, in all practical cases the relaxation time is usually much longer than any operational time; hence, the sequential device can be considered a system that remembers the last transition. Examples

include electronic flip-flop and DRAM (dynamic random access memory): the complex ‘storage capacitor + transistor’. They are employed in computers to perform the role of registers and memory cells. A simple mechanical example of sequential binary switch is the switch illustrated in **Figure 6**.



**Figure 6.** Mechanical binary switch as a sequential device: when an external force is removed it stays in the last state for any interval of time, shorter than the relaxation time.

In order to discuss the energetic behaviour of the two classes of binary switches, we need to introduce a dynamical model that is capable of representing the action of the force and the switch mechanism. In order to do so, we use a simple model based on a single degree of freedom  $x(t)$  that represents the system state (this can be the position of a pebble or the value of some electric voltage or current, as we discussed above). This quantity  $x(t)$  must be subjected to constraints and forces that make it to behave according the two rules (1) and (2) and also some time evolution equation, according to physics.

For both classes of devices, we can use the following equation:

$$m\ddot{x} = -\frac{d}{dx}U(x) - m\gamma\dot{x} + \xi(t) + F \quad (13)$$

where  $m$  represents the inertia of our system,  $F$  is an external force that can be applied when we want to change state, and  $\gamma$  is the frictional force that represent dissipative effects in the switch dynamics and

$$U(x) = \frac{1}{2}x^2 \quad \text{for combinational devices} \quad (14)$$

$$U(x) = -\frac{1}{2}x^2 + \frac{1}{4}x^4 \quad \text{for sequential devices} \quad (15)$$

In general,  $U(x)$  is a potential function that has the role of confining  $x(t)$  in a well-defined region. What is  $\xi(t)$ ? This is a stochastic force and represents the role of fluctuations that are unavoidably present due to a finite temperature. These fluctuations are responsible for the relaxation process that we discussed above. In a macroscopic binary switch, this term is quite small compared to the other terms in the equation of motion and is usually neglected. However, when we deal with micro- to nanoscale devices, like in modern binary switches, its role might be relevant and its presence cannot be neglected [2].

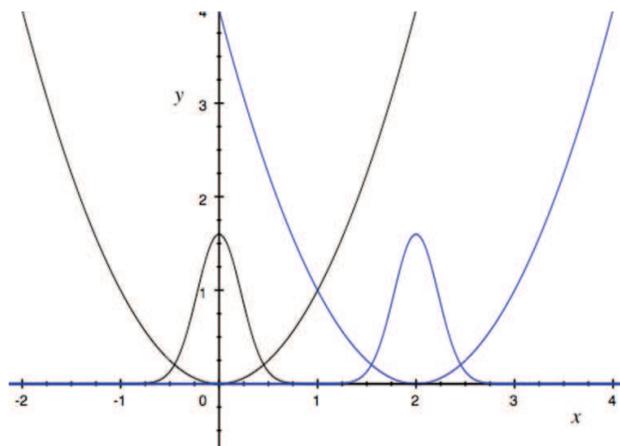
The fluctuating force  $\xi(t)$  is represented here by a zero average stochastic process that is defined in statistical terms. Due to its presence, the equation of motion is a stochastic equation

and its solution is usually described in statistical terms.  $P(x, t)dx$  represents the probability for the quantity  $x$  to be at time  $t$  within the interval between  $x$  and  $x + dx$  and is a relevant quantity to describe the system dynamics.

Here, we can define the two distinguishable physical states  $S_0$  and  $S_1$ , as follows: the state  $S_0$  is realized when  $x < x_{\text{TH}}$ ; the state  $S_1$  is realized when  $x > x_{\text{TH}}$ , and  $x_{\text{TH}}$  is a value of the quantity  $x$  that can be chosen conveniently. Due to the presence of fluctuations, the two physical states are assumed with a certain probability given by:

$$p_0(t) = \int_{-\infty}^{x_{\text{TH}}} P(x, t) dx \text{ and } p_1(t) = \int_{x_{\text{TH}}}^{+\infty} P(x, t) dx \quad (16)$$

The switch event in a combinational device is illustrated in **Figure 7**.



**Figure 7.** Switch event for the combinational device: the application of a constant force,  $F_{01} = -F_0$ , produces a net displacement of the  $p(x)$  from the one in black (centered around  $x = 0$ ) and corresponding to the state  $S_0$  to the one in blue (centered around  $x = 2$ ) and corresponding to the state  $S_1$ .

Here, the application of a constant force,  $F_{01} = -F_0$ , produces a net displacement of the  $p(x)$ . By setting properly the value of the threshold, we can easily realize the switch from  $S_0$  to  $S_1$ . According to the combinational character of our device, once the force is removed, the system reverts back to the initial state  $S_0$ . In order to compute the amount of energy required for this switch, we should take into account the work done by the forces acting on the system. The stochastic force does not do (on average) any work because it is a zero mean force. The dissipative force does a negative work that is proportional to the switch speed. The external force acts on the potential and is a conservative force, thus through an entire cycle its work is null.

The role of the dissipative force and of the fluctuating force can be, more properly, discussed within the thermodynamics framework that we have previously introduced. In fact, their presence accounts for the existence of a coupling of our quantity  $x(t)$  with a thermal bath that is responsible at the same time for the fluctuating part of the dynamics (i.e. the random force  $\xi(t)$ ) and the dissipative part (i.e. the damping constant  $\gamma$ ). Indeed the two are connected through a famous relation called the fluctuation-dissipation theorem [1] established by Harry

Theodor Nyquist (1889–1976) in 1928, and demonstrated by Callen and Welton in 1951. This relation is:

$$G_R = \frac{mK_B T}{\pi} \gamma \quad (17)$$

where  $G_R$  represents the intensity of the fluctuation with white noise spectrum [1].

Due to the existence of the thermal bath, thermodynamics sets the rule for the energy balance during the switch process. Specifically, if we conduct the switch process from the initial state  $S_0$  to the final state  $S_1$  we need to spend a minimum energy (i.e. producing a heat  $Q$  that goes into the thermal bath). In general, we have

$$TdS = dQ + E_d \quad (18)$$

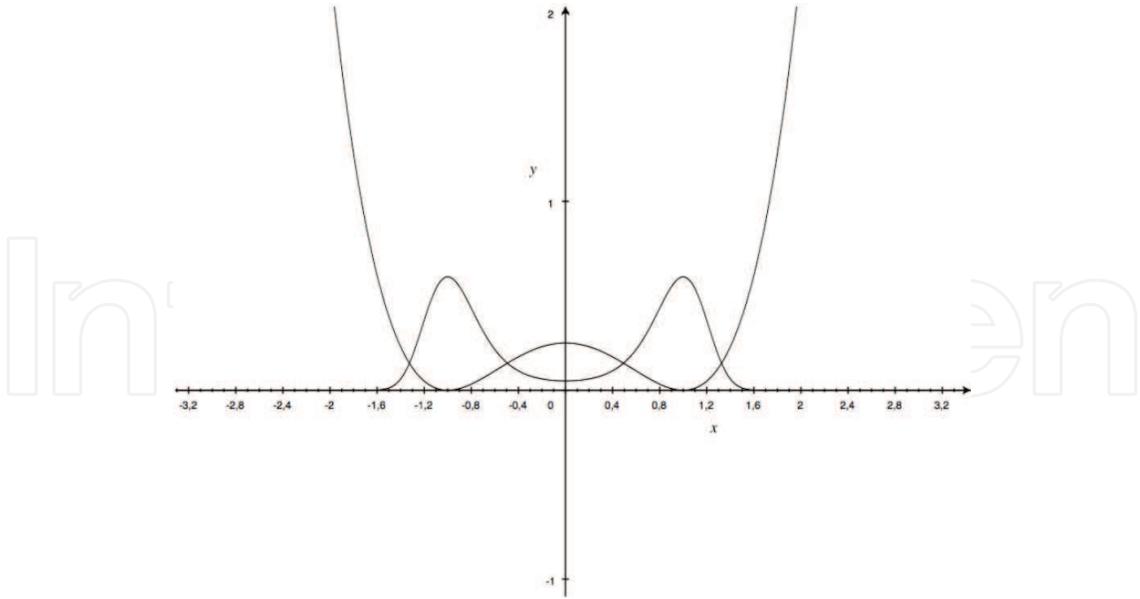
If the transformation is carried out in a reversible manner,  $E_d = 0$  and the amount of dissipated heat is, according to Clausius,  $dQ = TdS$ . In a cyclic operation, this is clearly zero because  $T$  is constant and  $S$  is a function of state only. On the other hand, if the transformation is not reversible, the amount of dissipated energy is  $E_d$  and is larger than zero.

Could this be realized in practice? In a recent experiment, Lopez-Suarez and et al. [3] has built a micro-cantilever that can be operated as a combinational device. They showed that by slowing down the switching operation,  $E_d$  can be made arbitrarily small, thus confirming that the minimum energy required to process information with a combinational device is indeed zero.

Let us now consider the switch event in a *sequential device*.

For this case [1], the definition of the switch event itself must be reconsidered. Previously, the switch event was defined as the change from an equilibrium position (e.g. at rest at the bottom of the potential well) to another equilibrium position (e.g. at rest at the bottom of the displaced potential well). In this bistable potential, however, the particle is never at rest at the bottom of a single well: due to the presence of the fluctuating force, the particle will be randomly oscillating around the potential minima, with occasional jumps between the two wells. Since the potential is symmetrical and we have a zero-mean fluctuating force, the two states  $S_0$  and  $S_1$  have the same probability. This implies that the probability density distribution at equilibrium  $P(x,t) = P(x)$  is stationary and symmetric, as represented in **Figure 8**.

When the particle is initially at rest at the bottom of the left well, after some time  $\tau_1$  it starts to oscillate around the potential minima and after some longer time  $\tau_2$  it will jump into the right well and eventually back into the left well and so on. The time  $\tau_1$  and  $\tau_2$  are random variables. Their mean values  $t_1 = \langle \tau_1 \rangle$  and  $t_2 = \langle \tau_2 \rangle$  (with  $t_2 > t_1$ ) can be computed on the bases of the features of the potential  $U(x)$  and the stochastic force  $\xi(t)$ . They are usually addressed as the *intra-well* relaxation time and the *inter-well* relaxation time and, in general, they represent, respectively, the average time the system takes to establish equilibrium within one well and the average time it takes to go to global equilibrium. Since  $t_2$  depends exponentially on the barrier height between the two wells, in practical switches the barrier height is chosen to be large enough to guarantee that  $t_2 \gg t_1$ .



**Figure 8.** Bistable potential  $U(x)$  with superimposed the probability distribution  $P(x,t) = P(x)$  at equilibrium.

Based on these considerations, we can define the switch event as the transition from an initial condition towards a final condition, where the initial condition is defined as  $\langle x \rangle < 0$  and the final condition is defined as  $\langle x \rangle > 0$ . With the initial condition characterized by:

$$p_0(t) = \int_{-\infty}^0 P(x, t) dx \cong 1 \text{ and } p_1(t) = \int_0^{+\infty} P(x, t) dx \cong 0 \tag{19}$$

and the final condition by

$$p_0(t) = \int_{-\infty}^0 P(x, t) dx \cong 0 \text{ and } p_1(t) = \int_0^{+\infty} P(x, t) dx \cong 1 \tag{20}$$

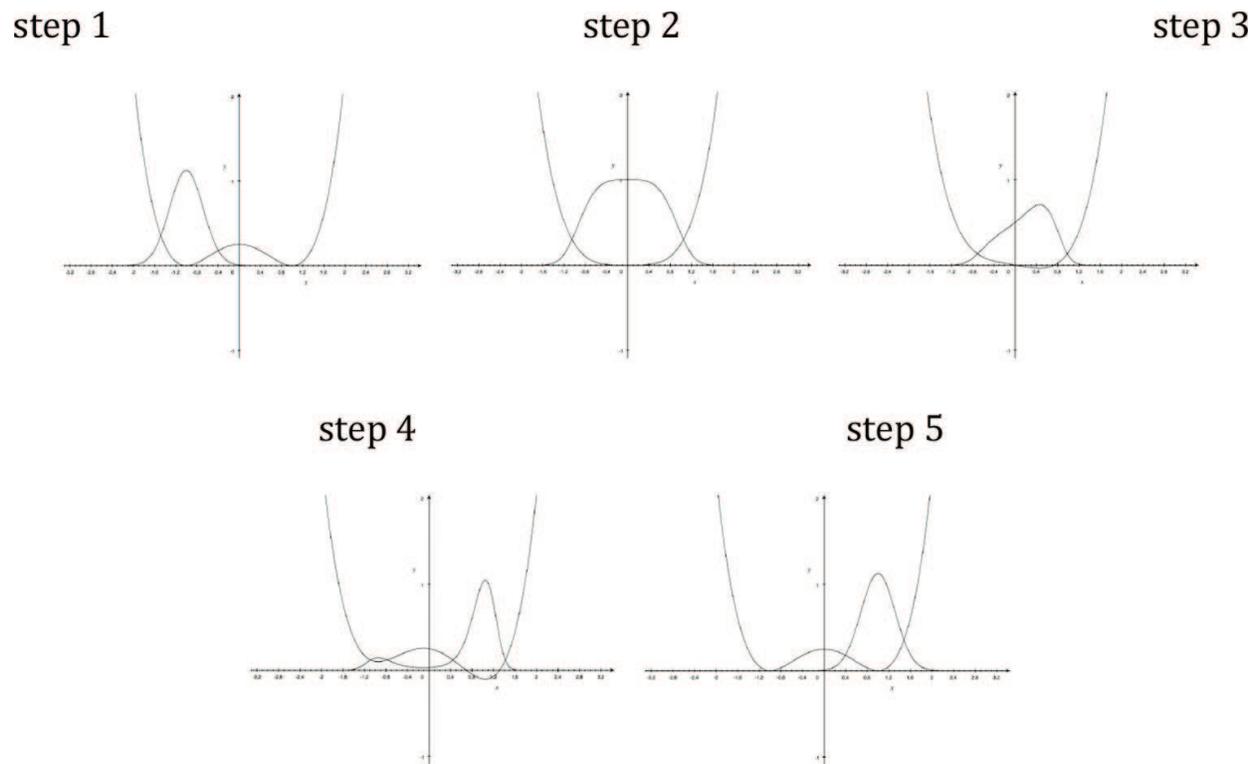
The conditions are reversed for a switch event from  $S_1$  to  $S_0$ .

In order to produce the switch event, we proceed as follows: set the initial position at any value  $x < 0$  and wait a time  $t_a$ , with  $t_1 \ll t_a \ll t_2$ , then apply an external force  $F$  for an elapsed time  $t_b$  to produce a change in the  $\langle x \rangle$  value from  $\langle x \rangle < 0$  to  $\langle x \rangle > 0$ . Then remove the force. In practice, it will be necessary to wait a time  $t_a$  after the force removal in order to verify that the switch event has occurred, i.e. that  $\langle x \rangle > 0$ . The total time spent has to satisfy the condition  $2 t_a + t_b \ll t_2$ .

Now that a switch event has been defined in this new framework, we can return to the question: what is the minimum energy required to produce a switch event?

As before, for the combinational device, it is quite easy to see that in order to minimize the energy dissipation, the role of the friction has to be negligible. In addition to this condition, we need to make sure that during the transformation no irreversible increase of the entropy takes place. The most common case (to be avoided) is the free expansion. During a free expansion, the system does not do any work and the entropy increases without energy expenditure. However, when we

need to bring back the system to its original state we cannot perform the reverse operation without energy expenditure because we need to decrease the entropy and this cannot be performed for free. This condition is particularly relevant for a procedure that is often followed in the switch event. The procedure is shown in **Figure 9** and consists in five subsequent steps.



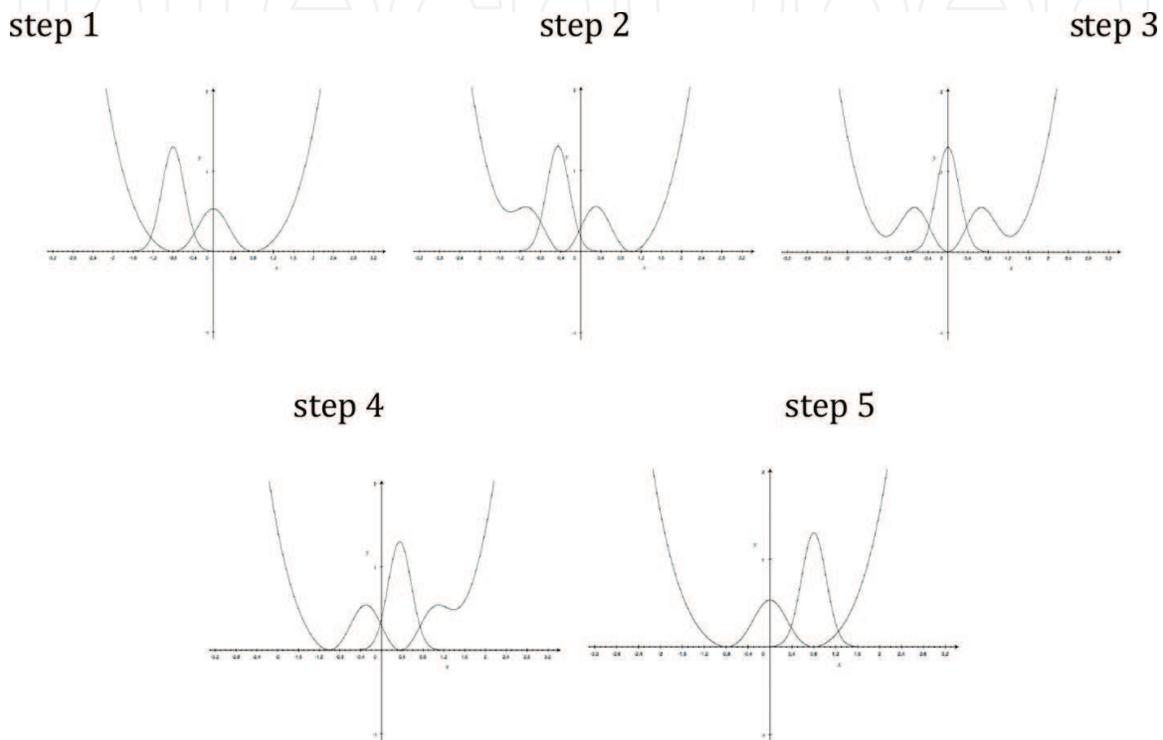
**Figure 9.** Potential  $U(x) + F$ . Equilibrium  $P(x)$  for the different cases. Second procedure: Step 1 and Step 5,  $F = 0$ ; step 2  $F = -x$ ; Step 3  $F = -x - F_1$ ; Step 4  $F = -F_1$ .

We point out that from step 1 to step 2, the entropy of the system increases. During this transformation, the potential changes by lowering the barrier. At this point, the particle dynamics relaxes (in a very short time) to the new configuration and the entropy increases like in a free expansion. This is apparent by the change in the probability distribution and can be demonstrated by simply assuming that in step 1 we have  $p_0 = 1$  and  $p_1 = 0$ , this gives  $S_1 = -k_B \ln 1 = 0$ . In step 2,  $p_0 = p_1 = 1/2$  and thus  $S_2 = -k_B (\frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \frac{1}{2}) = k_B \ln 2$ . Thus,  $\Delta S = k_B \ln 2 > 0$ . On the other hand, when there is a transition from step 2 to step 5, the entropy is reduced from  $S_2$  to  $S_5 = S_1 = 0$ , thus  $\Delta S = -k_B \ln 2 < 0$ . According to the thermodynamics, these last steps cannot be performed without providing energy to the system and thus the minimum energy in this case is not zero [1].

Summarising, the conditions required to perform a switching event that takes zero energy, are: (1) the total work performed on the system by the external force has to be zero. (2) The switch event has to proceed with a speed arbitrarily small in order to have arbitrarily small losses due to friction. (3) No free expansion entropy increase during the procedure.

In the following, we show a procedure (called *zero-power* protocol in the literature [4]) that satisfies these three conditions. In order to satisfy condition (1), we apply a force that maintains

the average position of the particle always close to the minimum of the potential well. In this case, the force is zero and thus the work is zero. In order to satisfy condition (2), we apply very slowly a change in the force. Finally, in order to satisfy condition (3), i.e. the probability density in state 0 and in state 1 is the same; apply a force that does not change the probability density along the path (constant entropy transformation). This can be done by applying a force that alters the potential, as shown in **Figure 10**. Such a procedure clearly satisfies the three conditions that we enunciated above.



**Figure 10.** Potential  $U(x) + F$ . Equilibrium  $P(x)$  for the different cases. Third procedure.

According to the reasoning developed in this procedure, we can conclude that also in the case of sequential devices, the minimum energy required to process information, i.e. to operate a switch, is indeed zero.

## 5. Energy efficiency, Landauer reset and reversible computing

In the previous sections, we have seen that a generic computing device can be considered as a machine that processes information while transforming work into heat. Pioneering research developed by J. Von Neumann and by R. Landauer in the last century has shown that information processing is intimately related to energy management ('information is physical' [5]). As a matter of fact, an ICT device is a machine that inputs information and energy (under the form of work), processes both and outputs information and energy (under the form of heat).

According to this model, energy efficiency during computation can be defined in terms of the quantity of input energy that is used for computation against the quantity that is transformed into heat. Thus, we can define energy efficiency as:

$$\eta = \frac{L-Q}{L} \quad (21)$$

$L$  is the input energy (in the form of work) and  $Q$  is the wasted heat produced during computation. Clearly, it varies between 0 (a totally inefficient device: all the input energy is wasted into heat) and 1 (maximum efficiency where all the energy is used to perform computation and none is wasted into heat).

Based on this definition, it is clear that the effort to reach the maximum efficiency is equivalent to the effort to reach the zero heat produced condition  $Q = 0$ . We note that, in principle, a computing device can be operated by keeping, during computation, the total change in the internal energy  $U = L - Q = 0$ . So the minimum energy required is  $L = Q$ .

The question that we aim to address since the beginning of the chapter is the following: is there a limit to how small can we make  $Q$ ?

This topic has been widely discussed in the scientific community since the beginning of the modern computers era. Based on our previous discussion, our question, by the moment that all the combinational and sequential devices can be made by interconnecting respective binary switches, translates into:

1. What is the minimum amount of energy required to operate a combinational switch?
2. What is the minimum amount of energy required to operate a sequential switch?

Based on the reasoning developed in the previous paragraph, we can now summarize the answer.

The answer to question (1) is  $Q = 0$ , provided that the switch operation is performed slowly enough in order to make negligible all the dissipative phenomena such as viscous damping and internal friction. Sometime this way of operating switches is called *adiabatic computing*.

The answer to question (2) is  $Q = 0$ , provided that the switch operation is performed slowly enough and that there is no irreversible entropy increase during the process (and the necessary and costly subsequent entropy decrease).

While the adiabatic computing condition is common to both classes of switches, the sequential devices demand an additional condition that requires some further discussion.

In fact, there exists a situation where the second condition in (2) cannot be satisfied. It is when the sequential switch is relaxed to equilibrium and the initial state condition is shared with equal probability by  $S_0$  and  $S_1$ . In this case, it is common to say that 'the knowledge of the system state is lost'. In order to apply the *zero-power* protocol it is necessary to put the system into a known state with 100% probability. This operation is called *Landauer reset* and cannot be performed without entropy reduction [5]. By the moment that on average, it requires a

reduction of the number of initial states from 2 to 1, the entropy decreases for a quantity  $\Delta S = K_B \text{Log } 2$  (it halves the state space) and this is necessarily associated with an amount of minimum energy to be dissipated  $Q = K_B T \text{Log } 2$ .

In conclusion, we have shown that a computing device can be operated with arbitrarily low energy expenditure, provided that no *Landauer reset* is required. Otherwise, a minimum energy expenditure has to be accounted in the measure of  $K_B T \text{Log } 2$  per reset.

Needless to say that this conclusion regards what we have called the 'fundamental limit' in energy consumption, during computation. Clearly, other limits [1] can arise when we deal with practical realization of computing devices. However, even if these limits are presently much larger and more important for practical applications, we should not forget that they are associated with the specific technology used. By changing technology, we can (in principle) always aim at reaching the fundamental limits.

## 6. Energy bounds on communication as information transfer processes

In this section, we discuss the concept of information transmission and its implication on the amount of energy required. This topic has been addressed since the beginning of last century and has been put in the modern form by the father of information theory, Claude Shannon.

The starting point of Shannon's reasoning is the following: if we want to transmit a certain amount of information (i.e. a message) through a given communication channel (it could be air, vacuum, copper wire, etc.) we want that this information reaches its destination uncorrupted. The cause of potential corruption is called noise. Shannon was able to demonstrate that if in a given channel characterized by a bandwidth  $B$ , at the destination you can measure an amount of noise power  $N$  and a signal power  $S$ , then the maximum amount of information per unit time (bit per second) you can transmit (without corruption) is:

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) \quad (22)$$

This relation is often addressed as the Shannon-Hartley theorem. We are interested in finding the minimum energy  $E_b$  that is required to transmit a single bit through a channel with a certain amount of noise. In order to find  $E_b$ , we need to express it in terms of the quantities in the Shannon's relation. By definition, the energy per bit is equal to the signal power  $S$  (energy per unit time) divided per capacity  $C$  (bit per second):  $E_b = S/C$ . On the other hand, the noise power  $N$  is equal to the noise spectral density  $N_0$  (noise power per unitary bandwidth) times the bandwidth  $B$ :  $N = N_0 B$ .

Thus, the previous relation becomes:

$$C_B = \log_2 \left( 1 + \frac{E_b}{N_0} C_B \right) \quad (23)$$

where we have introduced the quantity  $C_B = C/B$ , capacity per unitary bandwidth. The quantity  $E_b$  is readily obtained as:

$$E_b = \frac{(2^{C_B}-1)}{C_B} N_0 \quad (24)$$

If the channel bandwidth is much larger than the capacity (meaning that we transfer bits very slowly), we can take the limit when  $C_B$  goes to zero. By the moment that

$$\lim_{x \rightarrow 0} \frac{(a^x-1)}{x} = \ln a \quad (25)$$

we have  $E_b = N_0 \ln 2$ . The role of noise here is quite evident. How large it could be? Well, in principle, we can try to reduce the noise as much as possible but, even if we are able to suppress every other noise source, there is one source that cannot be avoided, which is the thermal noise, it is present in every physical system that is at finite temperature  $T$  and represent the natural oscillations of their elementary components (atoms and molecules). For the thermal noise, we have  $N_0 = K_B T$ . Thus, the minimum energy required for sending a bit of information, due to the fundamental noise limit, is

$$E_b = k_B T \ln 2 \quad (26)$$

As before, we would like to stress that this is fundamental limit that sets the minimum energy required. In practical systems, there are other noise sources that can play a relevant role as well.

One final word should be spent to compare this fundamental limit, i.e. the minimum energy required to transmit one bit, with the minimum energy required to do a switch event, i.e. the elementary step in the computation process. As we have seen that there is a minimum energy to be spent only in the case in which a Landauer reset is required. This amount of energy is dissipated in heat and it is definitively lost during information processing. For the minimum energy required to send one bit, instead,  $E_b$  is the energy associated with the physical signal that transmits the information. This quantity is not automatically dissipated in heat and, in principle, could be restored once the signal is received at the destination.

## Author details

Luca Gammaitoni

Address all correspondence to: [luca.gammaitoni@nipslab.org](mailto:luca.gammaitoni@nipslab.org)

NiPS Laboratory, University of Perugia, Perugia, Italy

## References

- [1] Gammaitoni L. Energy management at the nanoscale. (In ICT - Energy - Concepts Towards Zero - Power Information and Communication Technology, G. Fagas (Ed.), InTech (2014). DOI: 10.5772/57345. Available from: <http://www.intechopen.com/books/>

ict-energy-concepts-towards-zero-power-information-and-communication-technology/  
energy-management-at-the-nanoscale.

- [2] See the chapter Zhirnov V, Cavin R, Gammaitoni L. Minimum energy of computing, fundamental considerations. In ICT - Energy - Concepts Towards Zero - Power Information and Communication Technology, InTech [1].
- [3] Lopez-Suarez M, Neri I, Gammaitoni L. Sub  $k_B T$  micro-electromechanical irreversible logic gate, *Nature Communications* 7,12068 (2016) doi:10.1038/ncomms12068.
- [4] Gammaitoni L, Chiuchiú, Madami M, Carlotti G. Towards zero-power ICT. *Nanotechnology*, 2015 Jun 5; 26(22):222001.
- [5] Landauer R. Irreversibility and heat generation in the computing process, *IBM Journal of Research and Development* 1961; (5):183–191.

### Interesting textbooks:

Information Theory: A Tutorial Introduction, James V Stone, Paperback – February 1, 2015.

On-line textbook: Information Theory, Inference, and Learning Algorithms, by David MacKay (2005). (available here: <http://www.inference.phy.cam.ac.uk/mackay/itila/>)

IntechOpen