Corrigendum: Towards zero-power ICT (2015 Nanotechnology 26 222001)

Luca Gammaitoni¹, D Chiuchiù¹, M Madami² and G Carlotti²

¹NiPS Laboratory, Dipartimento di Fisica, Università di Perugia Italy
²Dipartimento di Fisica, Università di Perugia Italy

E-mail: luca.gammaitoni@nipslab.org

Received 31 May 2016
Accepted for publication 31 May 2016
Published 21 June 2016

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

The inset in figure 12 is incorrect and the considerations presented in 6.2. Energy versus error in the switch operation may be misleading for judging the impact of the error on the total energy dissipation. The correct inset picture is illustrated in figure 1 (below). Minimum energy dissipation for the reset-and-switch operation $Q_T(P_e, P_s) = Q(P_e) + Q_s(P_s, P_s)$. Where $Q(P_e)$ is the minimum energy dissipation in the reset operation, occurring with error probability $P_e$ as in equation (8) in the article (and reported here for convenience):

$$Q(P_e) = k_B T (1 - P_e) \ln(1 - P_e) + P_e \ln(P_e)) + k_B T \ln(2).$$

$Q_s(P_s, P_s) = Q(P_s) - Q(P_s)$ is the minimum energy dissipation in the subsequent switch operation, by assuming that it occurs with error probability $P_s$.

We acknowledge useful discussion with Igor Neri.

![Figure 1](image-url)

Figure 1. Grey: minimum energy dissipation $Q_T(P_e, P_s) = Q(P_e) + Q_s(P_s, P_s)$ for the reset-and-switch operation as a function of the error probability $P_e$. Here $P_s = 0.2$. Blue: $Q(P_e)$, minimum energy dissipated due to the initial reset occurring with error probability $P_e$. Black: $Q_s(P_s, 0.2)$, minimum energy dissipated due to the subsequent switch, occurring with switch error probability $P_s = 0.2$. 
Towards zero-power ICT

Luca Gammaitoni\textsuperscript{1}, D Chiuchiú\textsuperscript{2}, M Madami\textsuperscript{3} and G Carlotti\textsuperscript{3}

\textsuperscript{1}NiPS Laboratory, Dipartimento di Fisica, Università di Perugia and INFN Perugia, Via A. Pascoli 1, 06100 Perugia, Italy
\textsuperscript{2}NiPS Laboratory, Dipartimento di Fisica, Università di Perugia, Perugia, Italy
\textsuperscript{3}Dipartimento di Fisica, Università di Perugia, Perugia, Italy

E-mail: luca.gammaitoni@nipslab.org

Received 8 January 2015, revised 23 March 2015
Accepted for publication 26 March 2015
Published 11 May 2015

Abstract

Is it possible to operate a computing device with zero energy expenditure? This question, once considered just an academic dilemma, has recently become strategic for the future of information and communication technology. In fact, in the last forty years the semiconductor industry has been driven by its ability to scale down the size of the complementary metal-oxide semiconductor-field-effect transistor, the building block of present computing devices, and to increase computing capability density up to a point where the power dissipated in heat during computation has become a serious limitation. To overcome such a limitation, since 2004 the Nanoelectronics Research Initiative has launched a grand challenge to address the fundamental limits of the physics of switches. In Europe, the European Commission has recently funded a set of projects with the aim of minimizing the energy consumption of computing. In this article we briefly review state-of-the-art zero-power computing, with special attention paid to the aspects of energy dissipation at the micro- and nanoscales.

Keywords: ICT, energy, logic switch, landauer, computing

1. Introduction

In recent years, we have assisted the exponential growth of the information and communication technology (ICT) sector, fostered mainly by the progress made in the semiconductor industry in cost-effectively scaling down the size of solid-state electronic devices. Chief among all of this progress is the decrease in the size of electronic switches (i.e., transistors, complementary metal oxide semiconductor-field-effect transistor (CMOS-FET))\textsuperscript{4}, which are the building blocks of today’s computing devices. This decrease in size has been paralleled by the increase in computing capability due to ever-increasing transistor density, meaning that there is an ever-increasing number of computing elements in each device. However, this increase in density produced an increase in the power of the in heat dissipated during computation, up to the point that this is currently a serious practical limitation [1, 2]. According to the International Technology Roadmap for Semiconductors (ITRS) [3] the limits imposed by the energy dissipated during switch operation will be a roadblock for future scaling over the next 10–15 years. As a consequence, increasing attention has been devoted to the role of energy efficiency in ICT, and a new field called toward zero-power computing has emerged.

The objective of the research in this field covers both scientific and technological aspects. On the scientific side, there is a need to better understand the fundamental physical laws that rule the out-of-equilibrium energy transformation processes at the micro and nanoscales. On the technology
side, there is the goal of identifying new solutions for information processing that present improved energy efficiency and allow the continuous growth of computing capabilities.

1.1. Science aspects

The study of energy transformation processes at the micro- and nanoscales raises the problem of evaluating the efficiency of such a transformation. This is a topic that has been the focus of a significant research effort during the foundation of thermodynamics in the 1700s and 1800s. Thanks to the work of scientists like Sadi Carnot and subsequently of Emile Clapeyron, Rudolf Clausius, and William Thomson (Lord Kelvin), studies on the efficiency of machines invented by Thomas Newcomen and James Watt to transform heat into work brought us the notion of entropy and the second law of thermodynamics, which put limits on this efficiency. It is interesting to note that today, 200 years after the work of Carnot, we are still dealing with the problem of defining such limits, although today the object of our interest has moved from the large power plants of the industrial revolution to the tiny devices of modern ICT [4]. It is our understanding that future ICT will be largely affected by nanoscale devices that process information while transforming work into heat and heat into work. As a consequence, it seems natural to treat an ICT device as a novel infothermal machine that inputs information and energy (in the form of work), processes both, and outputs information and energy (in the form of heat).

1.2. Technology aspects

On the technology side, other than the urgent objective of reducing the heat produced during computation in high-performance computer systems, there is also another reason why zero-power computing is relevant for ICT. It is associated with the popular concept of ubiquitous computing. The introduction of microelectronic and mechanical systems and nanoelectronic and mechanical systems opened the door to the development of so-called wireless sensors: submillimeter sensors, computers, and actuators that can be interconnected to form a wireless network. These networked devices have wide applications that cover everything from military to civilian applications such as environmental monitoring, biomedical sensing, radio-frequency identification (RFID), interactive control, integrated biology, agriculture, structural monitoring, sensing harmful chemical agents, location of persons, and transportation, just to name a few.

The low-power characteristics of wireless sensor network components and the design of the system architecture are crucial to the longevity of the sensor nodes. The best solution for avoiding the impractical battery-replacement procedure is for each node to be autonomous and self-powered, adsorbing energy from a renewable source that is continuously available in the ambient environment. This approach is called energy harvesting. At present, renewable power can be obtained by generating electrical energy from light, thermal energy, and kinetic energy present within the sensors, environment. The energy from these sources can be used both as a real-time supply for the functioning of the device or can be stored for a later use to augment the battery, thereby increasing the lifetime and capability of the network and mitigating the environmental impact caused by the disposal of batteries. However, before the energy-harvesting technologies become viable solutions, it is necessary to reduce the gap between how much energy can be transformed from the environment in a certain time interval and how much power is required to operate the microdevice. Presently this gap spans about three orders of magnitude (see figure 1), with differences related to different energy-harvesting technologies. For a discussion of these aspects, see [5].

The energy consumption of computing technologies not only becomes more and more of an obstacle to realizing new functionalities in, for instance, mobile or distributed applications, and limits high-performance computing, it also has an increasing impact on energy supply and the environment. According to the SMART2020 study [6], "Enabling the low carbon economy in the information age, ICT’s share of global energy consumption is presently in the range of 2–5%. To reach the objective for CO2 emissions targets of at least 20% below 1990 levels in 2020, the SMART2020 report points at the strategic role played by ICT: ‘The ICT sector’s own emissions are expected to increase, in a business as usual (BAU) scenario, from 0.53 billion tonnes (Gt) carbon dioxide equivalent (CO2e) in 2002 to 1.43 GtCO2e in 2020. But specific ICT opportunities identified in this report can lead to emission reductions five times the size of the sector’s own footprint, up to 7.8 GtCO2e, or 15% of total BAU emissions by 2020’.

As illustrated in figure 2 [7], in modern electronic ICT systems, energy consumption is divided among the four main information processing functions: computation, communication, storage, and display. In the following we focus on the
role of computation, being the most basic and omnipresent function in practically all ICT devices.

2. Energy dissipation in existing nanoelectronic switches

The energetics of computation has attracted much interest in recent years due to the potential impact of this issue on the development of ICT. As we briefly mentioned in the introduction, the power dissipated during computation is a serious limit to the further increases in transistor density. Since 2004, the Nanoelectronics Research Initiative (nri.src.org) was formed to address the fundamental limits of the physics of logic switches, the basic element of the modern computer. Such an initiative has been somehow paralleled in Europe by Aeneas with the so-called More than Moore approach and recently with the FET Proactive initiative denominated MINECC (Minimizing Energy Consumption of Computing to the Limit)\(^5\).

The information developed in this framework focused mainly on the operation of an FET transistor, which is used here as a logic switch. The progressive miniaturization of FET transistors has permitted larger and larger densities, up to the point that the present generation of commercial microprocessors presents a density as high as 2 billion transistors per square centimeter [8]. Such a high transistor density of transistors is responsible for the phenomenon of device heating, which is presently the main technological limitation to a further increase in transistor density. In fact, in today’s microprocessors the power dissipated in heat is close to 100 W cm\(^{-2}\). It has been estimated that if the transistor density is exploited to the physical limits, the resulting power density for these switches at maximum packing density would be on the order of 1 MW cm\(^{-2}\) which is orders of magnitude higher than the practical air-cooling limit [2].

To better understand where and when the energy gets dissipated during computation we need to consider that the figures we discussed above are specific to existing electronic technology, called CMOS, and that they are mainly due to the peculiar operation procedure assumed in the functioning of the transistor itself. In fact, the main sources of energy consumption in electronics are the charging and discharging of electrical capacitances, which are present in all electronic devices and are associated with the functioning of the transistor as a logic switch. By assembling logic switches, we can build logic gates and memory devices and use them to perform all the logic and arithmetic operations required. Logic switches can also be used as a building block of memory devices.

In figure 3 we show an possible configuration where an FET is coupled to an external capacitor to form a dynamic random-access memory device. In this case, in addition to the gate capacitance $C_g$ of the FET and the interconnection capacitance, $C_i$, arising from the connection between different devices, figure 3 also shows the capacitance due to the storage capacitor, $C_s$. During the switch operation in an FET transistor, the capacitor is charged/discharged from/to a constant voltage power supply. This produces a total energy dissipated (i.e., converted into heat), per cycle, in the measure of [9]:

$$E_{sw} = CV^2$$

where $V$ is the voltage of the power supply. If the switching operation is performed with frequency $f$, then the power dissipated is:

$$P_{sw} = \alpha f CV^2$$

where $\alpha$ is the activity factor ($\alpha = 1$, for a square-wave switching). $P_{sw}$ is usually called the dynamic switching power and is directly associated with on/off switching (usually, in ordinary devices, there is also a parasitic leakage power component that adds to dissipation). In the last 40 years

\[\text{Figure 2. In modern electronic ICT systems the energy consumption is divided among the four main information processing functions: computing, communication, storage, and display.} \]

\[\text{Figure 3. Example of a dynamic random-access memory (DRAM) where several distinct capacitors are present: the gate capacitance, } C_g \text{ of the FET; the interconnection capacitance, } C_i \text{, arising from the connection between different devices; the capacitance, } C_s \text{ that results from to the storage capacitor.} \]
Dynamic switching power has been constantly reduced as a result of scaling down the devices’ critical dimension, $d$, from micrometers to nanometers. Such scaling has produced a proportional decrease in the device capacitance, $C_g$. The voltage, $V$, underwent a similar scaling down proportional to $d$. This situation has produced a progressive decrease in the dynamic switching power that is approximately proportional to $d^3$, a behavior confirmed by data obtained from several editions of the ITRS [3] and illustrated in [9].

It is interesting to note that while for devices with $d \approx 1 \mu m$, the switching energy of the transistor is the dominant factor in the total chip energy consumption, for sub-100 nm technology nodes, the fraction of the transistor dynamic energy in the energy balance decreases. Indeed, transistor dynamic energy consumption constitutes $\approx 20 - 30\%$ of the total energy in modern microprocessors (22–65 nm node), and is expected to further decrease for $d \approx 10$ nm devices and below [9]. This is mainly due to the increasingly important role of dissipation in interconnects and off-state leakage losses. The state-of-the-art switching energies with FET transistors are currently on the order of a few $10^{-18} J$.

In figure 4 we present the impressive decrease in energy consumption in switch devices used over computation in the last 80 years.

Due to the important role of switches in the future of ICT, it is relevant to address the following question: Is there any fundamental physical limit to the minimum energy required to operate a switch? Surprisingly enough, there is no general agreement on the answer to this question in the physics/technology community. To determine the correct answer to this question, we need to introduce a certain degree of abstraction to the treatment of the switch device. Let’s start by considering the role of a switch in a generic computation process. We can think of a switch as a generic physical system characterized by (at least) two stable states. Under the action of an external force, the system can change its state by switching from one state to the other. In figure 5 we present a schematic representation of a switch device where the external force is considered as input and the state (open or closed) is the output. In the electronic (or electro mechanical) realization of the switch, the status is represented by the value of an electrical current or an electrical voltage that changes as a consequence of the change in the state of the switch.

In this framework, a computation is performed every time a switch changes its status. Clearly by combining switches we can realize logic gates (and thus logic gate networks), memory devices, and perform all the computation required for a given task.

Let’s consider the potential energy that regulates the functioning of a switch, represented in figure 6. The two logic states, 0 and 1, are here represented as logic state 0, 1, associated with $x < 0$, $x > 0$. To be dynamically stable, the two states are separated by a potential energy barrier, $E_b$.
left and right well, respectively. The switching event is obtained by making the particle energetic enough to overcome the potential barrier separating the two states or, equivalently, by lowering the potential barrier on the particle side. Notwithstanding the simplicity of this model, it has often been applied [2] to estimate the relevant aspects of the physics of switches. Specifically, the minimum operational energy of the switch is computed by assuming that the barrier height, $E_b$, is large enough to enable the distinguishability of the two logic states. Such a condition is, in fact, threatened by unwanted crossings of the potential barrier due to thermally induced (classical) jumps or tunneling (quantum) effects. The larger $E_b$ and the distance between the two wells, the lower the threat to the distinguishability of the two states. Additionally, the Heisenberg energy-time uncertainty relation is invoked in this context to set a further limit to the barrier height, $E_b$. Based on these arguments, Cavin et al were able to estimate a minimum energy per switching event of approximately $3k_B T ln(2)$.

3. The physics of nanoscale switches

To continue in the search for the minimum energy required to operate a switch, we should introduce a general dynamical model that is capable of representing the main features of a physical switch like the one we have just described without being limited by the considerations associated with the peculiarities of the electric charge dynamics. Thus we propose to discuss the switch dynamics in terms of the continuous dynamics of a single degree of freedom, $x(t)$, ideally representing a relevant observable of the switch dynamics (e.g., the position of a material particle or of a cursor and the quantity of electric charge or magnetic or electric field). To represent the binary character of the switch, we need to identify the two logic states: we assume that the logic state 0 (1) is associated with $x < 0$ ($x > 0$). As we mentioned previously, the two states have to be dynamically stable and thus are separated by a potential energy barrier, $E_b$. Taking into account the physical character of a generic small-scale binary switch, we imagine it to be in contact with a thermal bath at a constant temperature, $T$.

The continuous dynamics of $x(t)$ can be mathematically described by a proper Langevin equation [11]:

$$m\ddot{x} = -\frac{dU_b(x)}{dx} - \gamma \dot{x} + \xi(t) - \frac{dU_i(x, t)}{dx}$$

where $U_b(x)$ is a symmetric stationary bistable potential and $f(x, t) = -dU_i(x, t)/dx$ is an external force that can be applied to change the logic state. $U(x, t) = U_b(x) + U_i(x, t)$ is the time-evolving potential energy landscape and $F(x, t) = -dU_b(x)/dx - dU_i(x, t)/dx$ is the total conservative force acting on the system. $\xi(t)$ represents the fluctuating force whose statistical features are connected with the dissipative properties, $\gamma$, by a proper fluctuation-dissipation relation [12]. The time evolution of the corresponding probability density, $p(x, t)$, is usually described in terms of the associated Fokker–Planck equation [11]. The two states, 0 and 1, are realized with the respective probabilities $p_0$ and $p_1$ ($p_0 + p_1 = 1$), given by:

$$p_0 = \int_{-\infty}^{0} p(x, t) dx, \quad p_1 = \int_{0}^{\infty} p(x, t) dx.$$  (4)

States 0 and 1 have the same energy thanks to potential symmetry, so to perform the switch operation with zero energy expenditure, we need to reduce to zero the work performed by the deterministic forces acting on the system. The deterministic forces are represented here by the conservative force $F(x, t)$ and the dissipative force $\gamma \dot{x}$. In the following, we exclude the case of a reversible transformation where the work performed against the conservative forces during the switch is stored as potential energy in some external place and recovered subsequently, because this condition is practically difficult to implement in a small-scale switch.

To begin our analysis, let’s start with the following procedure (see the left-hand side of figure 7, standard procedure) [13], which presents a switch from 0 to 1 (the 1 to 0 switching is analogous). We start with the system in the logic state, 0 (step 1: first picture from the top). Here the potential barrier, $E_b$, is
initially lowered to zero (step 2), and subsequently the potential is tilted in order to move the particle to the right well (step 3). Finally, the barrier is raised to its initial value (step 4) and the tilt is removed (step 5). All five steps in this procedure can be performed by a proper application of the external force, \( f(x, t) \), in order to keep the average velocity of the particle close to zero (zero friction) and the average position in correspondence to an approximately zero derivative potential value (zero deterministic force). However, in spite of negligible friction and negligible external work performed on the system, this procedure does not allow for a zero-power switching. This is due to the unavoidable entropy reduction occurring in steps 3–5, after its sudden increase in steps 1 and 2, due to the barrier drop. This is apparent from the change in the probability distribution (see figure 7, left) and can be demonstrated quantitatively as follows. The system entropy in the various steps can be computed according to Gibbs as:

\[
S_i = -k_B \sum_j p_i \ln p_i
\]

where we assume that the Gibbs entropy and the Shannon entropy coincide [19]. \( s = 0, \ldots, 5 \) indicates the step number and \( i = 0, 1 \). In step 1, we have \( p_0 = 1 \) and \( p_1 = 0 \), \( S_1 = 0 \). In step 2, \( p_0 = p_1 = 1/2 \) and thus \( S_2 = k_B \ln (2) \). Therefore, the change in entropy is \( \Delta S_{2-1} = k_B \ln (2) > 0 \). On the other hand, from steps 3–5, entropy is reduced from \( S_2 \) to \( S_5 \), \( S_5 = S_0 = 0 \), thus \( \Delta S_{5-2} = -k_B \ln (2) < 0 \). According to the thermodynamics, while the entropy increase can be performed without energy exchange, these last steps cannot be performed without energy expenditure. This result is also in agreement with what has been observed in the simulation of a nano-magnetic system [14] where this switch procedure has been applied.

Thus, based on this discussion, we are now in a position to answer the question that was posed at the beginning of this section. The minimum energy required to operate a switch is zero, provided that the following conditions are obeyed: (1) The application of the external force always keeps the average position of the particle close to the minimum of the potential well (zero total force). (2) The switch event has to proceed with as small a speed as possible (zero friction approximation). (3) The system entropy must remain constant during the switch event. The right-hand side of figure 7 shows an example of a possible procedure that satisfies these three conditions.

4. Nanomagnetic zero-power switches

To test these conclusions, a prototype nanoscale switch has been considered in the form of a single cylindrical element of permalloy (NiFe) with dimensions \( 50 \times 50 \times 5 \) nm\(^3\) discretized in elementary cells of \( 5 \times 5 \times 5 \) nm\(^3\). Micromagnetic simulations of the switch dynamics have been performed at \( T = 300 \) K, and the effect of thermal fluctuations is introduced into the simulation with a random fluctuating magnetic field, which is delta correlated both in time and in space, with an amplitude that fulfills the fluctuation-dissipation theorem [15].

To simulate a bistable system, we introduced a uniaxial anisotropy in the plane of the NiFe dot, along the \( y \)-direction. This anisotropy defines two low-energy states for the magnetization, \( +y \) and \( -y \), which can be defined as 0 and 1 states, respectively. These two energy minima are separated by an energy barrier whose height is determined by the value of the uniaxial anisotropy constant, \( K_y \).

The switching procedure was realized by applying a sequence of external magnetic fields, which is illustrated in the upper part of figure 8. (This procedure was inspired by the zero-power procedure previously described.) The presence of the uniaxial anisotropy along the \( y \)-direction is represented in the figure using an elliptical shape with its major axis parallel to \( y \). The system starts within no applied external field and with the initial magnetization oriented along the positive \( y \)-direction (i.e., the system is in one of the two energy minima). In the first stage, a positive external field (\( H \)) is applied along \( y \) with a slope, up to a maximum value of \( H = 1.0 \) kOe. After that, the external field is slowly rotated in the \( x-y \) plane of \( 180^\circ \) toward the negative \( y \)-direction, keeping its amplitude.
fixed at the value $H = 1.0$ kOe. During the final stage of the simulation, the external field is removed with a negative slope, opposite to that of the first stage. The energy dissipated during the switching procedure is then calculated as the integral of the scalar product $\mathbf{H} \cdot d\mathbf{M}$, where $\mathbf{M}$ is the average magnetization of the system. The total switch time was also varied in the range $2 \times 10^{-1} - 10^{0}$ ns. The results shown in figure 8 (lower) for the micromagnetic simulations represent the energy dissipation during the switching procedure as a function of the total switch time. The calculations were performed for a barrier height of about $40 k_B T$. In this switching procedure, the external field is applied parallel to the initial magnetization configuration so the system remains in the absolute minimum energy state during the whole procedure. For comparison, the same figure also shows the results obtained with the numerical solution of equation (1) [20].

Figure 8 also shows the results obtained with Euler–Maruyama integration of equation (3) with $m = 11 \times 10^{-15}$ Kg, $\gamma = 1.89 \times 10^{-10}$ Kg s$^{-1}$, $T = 300$ K. $\xi_\tau$ is an uncorrelated white noise, so its cumulants satisfy $\langle \xi_\tau \rangle = 0$ and $\langle \xi_\tau \xi_{\tau + \tau} \rangle = 2\xi K_T \delta(t)$. $U_0(x) = -ax^2/2 + bx^4/4$ with $a = 4.9680 \times 10^{-7}$ J m$^{-2}$ and $b = 4.9680 \times 10^{-5}$ J m$^{-4}$; $U_0(x)$ then has two minima in $x = \pm 1 \mu m$ separated by an energy barrier, $E_b = 30K_b T$. These choices are consistent with the dynamics of a one-dimensional micrometric bead dispersed in distilled water [13]. We assume that at $t = 0$, the bead is in local equilibrium within the left well of $U_0(x)$. To perform a zero-power switching, we then apply, for a time span of length $\tau$, an additional $U_1(x, t)$, such that the total potential energy $U(x, t)$ is a time-variable ninth-order polynomial. For large values of $\tau$, $U(x, t)$ gives one analytical representation of the zero power protocol shown in figure 7 (right). The average heat exchanged by the particle with the thermal bath is computed as the Stratonovic integral [20]

$$Q = \left\langle \int_{0}^{\tau} - \frac{dU(x, t)}{dx} \cdot dx - \int_{0}^{\tau} m \dot{x} \cdot dx \right\rangle,$$

where the angle brackets stand for the average over many simulated trajectories.

Both sets of data clearly show that the dissipated energy can be made as small as needed by increasing the total switch procedure duration, asymptotically reaching the zero energy dissipation limit, according to our prediction. Based on these considerations for the zero-power switching, we can critically address the recent observation [21] that in order to reach zero-power switch operation, a key element is represented by the existence of a copy of the switch already in the final destination status. We have shown here that this condition is not required in a binary switch. Moreover, we want to stress that the switch operation realized here is performed according to the leads discussed in the previous paragraph. Thus the transformation is reversible and does not decrease entropy. In the next section, we will consider a transformation where irreversibility via a net decrease of entropy arises: the so-called Landauer reset.

---

**Figure 9.** Potential energy $U(x)$ (blue) and equilibrium probability density $P(x, t) = P(x)$ (red).

---

5. Switch versus reset operation: the Landauer limit

In the dynamical description provided above (see (3)), we assumed that due to the coupling with the thermal bath, a fluctuating force, $\xi(t)$, appears. At thermal equilibrium, the fluctuation-dissipation theorem, [12] links $\xi(t)$ and the dissipative force, $\gamma$. In a generic computing device, the practical switch is supposed to rest in one of the two logic states between two subsequent switching procedures. In a nanoscale switch, the role of fluctuations cannot be neglected, and thus in our bistable model, due to the presence of the fluctuating force, the particle will oscillate around the potential minima, with occasional random crossings of the potential barrier between the two wells. More specifically, if we place the particle at rest at the bottom of the left well, it starts to oscillate, and after some time, $\tau_1$, it reaches a constant oscillation amplitude. If we wait long enough we can observe that after a time, $\tau_2$, it will jump into the right well and eventually back into the left well, and so on at a constant pace. $\tau_1$ and $\tau_2$ are random variables whose average values are usually referred to as intrawell relaxation time and the the interwell relaxation time, respectively. They represent the average time the system takes to establish local equilibrium within one well (as it would be if the potential was no wider than a single well) and the average time it takes to go to global equilibrium in the entire potential, respectively. By the moment that $\tau_2$ depends exponentially on the barrier height between the two wells, in practical switches, the barrier height is chosen to be large enough to guarantee $\tau_2 \gg \tau_1$.

In the description of the switch event, we implicitly assumed that the system is at local equilibrium in a given well and that the switching process proceeds within a time, $\Delta t$, such that $\tau_1 < \Delta t < \tau_2$. Now let’s suppose that instead, we are not at local equilibrium anymore but rather have reached a global equilibrium, (i.e., before starting the switch process, we have waited a time, $t > \tau_2$).
Since the potential is symmetrical and the fluctuating force has zero average, the two states, 0 and 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 9. Here, \( p_0 = p_1 = 1/2 \).

If we want to perform a switch operation, say 0 to 1 we need to operate a transformation that changes, the probability density distribution in order to reach the condition \( p_0 = 1 \), \( p_1 = 0 \). This transformation is usually called the reset. From this perspective, we can say that a switch operation transforms a local equilibrium condition into another local equilibrium condition, while a reset operation transforms a global equilibrium condition into a local equilibrium condition.

Similarly to what we have done for the switch operations, we can identify a proper procedure that minimize, the energy required to operate the reset by zeroing the work done by the deterministic forces. In this case, however, the resulting change in entropy cannot be made null. In fact, in the initial state we have \( p_0 = 1/2 \) and \( p_1 = 1/2 \), \( S_f = -k_B \ln (2) \), while in the final state we have \( p_0 = 1 \) and \( p_1 = 0 \), \( S_f = 0 \). Thus the change in entropy is \( \Delta S_{f\rightarrow f} = k_B \ln (2) > 0 \). According to the second principle of thermodynamics, this amounts to a minimum energy, \( Q \), to be dissipated:

\[
Q \geq k_B T \ln (2)
\]

where \( T \) is the temperature. This result, often called the Landauer principle, represents a version of the second principle of thermodynamics that, assuming the Shannon information as a special form of the Gibbs–Boltzmann entropy, establishes that a necessary condition to operate a computing device with zero energy dissipated is that the computing process does not decrease information [16–19]. It has recently been experimentally tested [13, 21, 22], with the aim of exploring the limits in low-power computation [14].

6. Operating nanoscale switches in the presence of fluctuations

When we operate the binary switch device in the presence of fluctuations—a condition that is typical in the case of nanoscale switches—we are faced with the problem of unwanted barrier-crossing events. In fact, even if we set our barrier high enough that our resulting \( \tau_2 \gg \Delta t \), there is always a finite probability that a jump occurs. The effect of these random crossings is interpreted as a switch error that produces a logical bit flip. Bit flips occur with a certain probability in any real switch, as a result of fluctuations in the presence of a finite barrier height [23]. In this section, we explore the role of bit-flip errors in determining the minimum energy.

6.1. Energy versus error in the reset operation

Let’s start with the reset operation. Let’s suppose that during the reset to 0 operation, a bit flip occurs with a certain probability, \( p_e = 1 - p_0 = p_1 \). As a result of the reset error, a finite probability, \( p_1 \), appears. In figure 10, we present the probability densities before and after the reset to 0.

To compute the minimum energy, let’s proceed with the computation of the entropy change. The Gibbs entropy before the reset operation \( p_0 = p_1 = 0.5 \) is promptly computed as \( S_f = k_B \ln (2) \). After the reset operation, we have:

\[
S_f(P_e) = -k_B \left( (1 - P_e) \ln (1 - P_e) + P_e \ln (P_e) \right)
\]

where we used \( p_0 = 1 - P_e \). Accordingly, the energy dissipated during the erasure operation is now a function of the error probability: \( Q(P_e) = T \Delta S(P_e) \)—that is,

\[
Q(P_e) = -k_B T \left( (1 - P_e) \ln (1 - P_e) + P_e \ln (P_e) \right) - k_B T \ln (2).
\]

In figure 11, we plot the energy ratio \( \eta = Q(P_e)/Q_0 \) as a function of the error probability, \( P_e \). As one can see, for \( P_e > 0 \) we have \( \eta < 1 \), implying that if we are willing to accept a larger-than-zero error probability, we can beat the Landauer limit and perform the resetting operation with an energy toll smaller than \( k_B T \ln (2) \). Consistently, the zero limit for energy dissipation is reached when \( P_e = 0.5 \), corresponding to the maximum uncertainty (complete indistinguishability) (i.e., no reset). In the \( P_e = 0 \) limit, we regain the Landauer’s prediction, \( Q(0) = Q_L \).
Figure 11. Energy ratio $\eta_{fi}$ as a function of the error probability, $P_e$. The energy ratio, $\eta_{fi}$, represents the fraction of Landauer’s energy required to perform the reset operation in the presence of a finite error probability, $P_e$. For a bistable latch ($N = 2$), if we accept an error probability of 20% ($P_e = 0.2$), we need to dissipate approx $1/4$ of the minimum required by the Landauer limit. Multistable devices ($N = 3, N = 4$) show an energy ratio, $\eta_{fi}$, larger than the bistable one. In the large $N$ limit, $\eta_{fi}$ becomes a linear function of $P_e$.

6.2. Energy versus error in the switch operation

Let’s consider the switch operation. We have already demonstrated that if we apply the zero-power procedure, we can perform the switch with zero energy expenditure. Now let’s consider the case in which we apply the 0-to-1 switch procedure when the state of the system, instead of being in the proper 0 initial state, is already in 1 state as a result of a bit-flip error. In this case, what we can demonstrate is that if after the procedure the final state is still 1, then the minimum energy cost of this operation is $2k_BT \ln (2)$.

This is easily demonstrated by assuming that we want to use this procedure for resetting to 1. We can compute the energy dissipated as the sum of two processes that are realized with probability 1/2 each: the switch 0-to-1 with energy $Q_A$ and the transformation 1-to-1 with energy $Q_B$. Due to the Landauer principle, the total energy dissipated is $Q = 1/2 Q_A + 1/2 Q_B \geq k_BT \ln (2)$. However, in hypothesis it was $Q_A = 0$ (because the minimum energy of this switch can be zero with a zero-power procedure); thus $Q_B \geq 2k_BT \ln (2)$.

In figure 12 we show the results of the digital simulation of equation (3), together with the micromagnetic simulations, as before. In this set of simulations, however, the initial magnetization of the system was set to be antiparallel to the applied external field, realizing process B. Here the system starts on a relative energy minimum, which becomes more and more unstable as the external magnetic field is increased in the opposite direction, until the magnetization is forced to reverse toward the absolute minimum energy state. Simulations show that the energy dissipated is much larger with respect to the process A, and it can be reduced toward the limit, $2k_BT \ln (2)$, while reducing the energy barrier height.

As we have anticipated, as a consequence of a bit-flip error, the switch is assumed to be in the wrong logical status; thus when we apply the proper zero-power procedure, we incur a minimum energy toll of $Q_{bf} = 2k_BT \ln (2)$, as just demonstrated. This argument is reminiscent of the minimum energy toll required when operating the electronic switches [2] but in contrast, this is of a fundamental nature and applies regardless to the technology employed of build the switch.

We observe that this result poses a limit to the trading between energy and uncertainty [24] in the reset operation we discussed above. As was observed [24], the Landauer limit can be beaten by accepting a finite error probability, $P_e$, during the reset. However, such a gain is now balanced by the additional energy dissipation due to the switch error condition that affects the subsequent switch. By putting together the two contributions, we can determine the minimum energy associated with the sequence reset-and-switch as a function of the reset error probability. The relation that generalizes the one obtained in [24] is expressed by $Q(P_e) = k_BT(1 - P_e) \ln (1 - P_e) + P_e \ln (P_e) + k_BT \ln (2) + P_e 2k_BT \ln (2)$.

In figure 12 (inset) we plot the minimum energy, $Q(P_e)$, as a function of the error probability, $P_e$. Remarkably, $Q(P_e)$ shows a minimum $Q(\bar{P}_e) = k_BT \ln (8/5)$, for $\bar{P}_e = 1/5$ independent from $T$. Such a behavior, characterized by the existence of an optimal error probability (e.g., due to a given noise intensity) that minimizes the reset-and-switch energy is reminiscent of a vast class of phenomena where the presence of a finite amount of fluctuations, instead of being detrimental, is actually beneficial [25].
7. Conclusions

We started this review by asking a provocative question: *is it possible to operate a computing device with zero energy expenditure?* Our question was motivated by the technological need to reduce the amount of energy currently dissipated during computation. To provide a meaningful answer, we addressed the most basic component of today’s computing devices: the logic switches. Based on our analysis of the physics of a general dynamical model of the switch, we concluded that, provided a proper switch protocol is applied, there is no lower bound to the minimum energy required to operate the switch. This result was tested with dynamical simulations of a nano-magnetic switch prototype. Further considerations on the functioning of the nano switches in the presence of large fluctuations have elucidated the role of noise-induced errors in switching energetics.

Acknowledgments

The authors gratefully acknowledge financial support from the European Commission (FPVII, Grant agreement no: 318287, LANDAUER and Grant agreement no: 611004, ICT-Energy), Fondazione Cassa di Risparmio di Perugia (Bando a tema Ricerca di Base 2013, Caratterizzazione e micro-caratterizzazione di circuiti MEMS per generazione di energia pulita) and ONRG grant N00014-11-1-0695.

References

[1] Bernstein K, Cavin R K, Porod W, Seabaugh A and Welser J 2010 *Proc. IEEE* **98** 2169–84
[2] Welser J J, Bourianoff G I, Zhirmov V V and Cavin R K III 2008 *J. Nanopart. Res.* **10** 1–10
[3] International Technology Roadmap for Semiconductors. www.itrs.net
[4] Pop E 2010 *Nano Res.* **3** 147169
[5] Gammaitoni L 2012 *Contemp. Phys.* **53** 119–35
[6] SMART 2020 Enabling the low-carbon economy in the information age, GeSI, 2008
[7] Nordman B and Lanzisera S 2011 Electronics and network energy use: Status and prospects *IEEE Int. Conf. on Consumer Electronics* 245–6
[8] Sebastian S 2014 The Apple A8 (dual-core ARM64 ‘mobile SoC’)) counts 2 billion transistors in 89 mm²
[9] Zhirmov L V, Cavin R and Gammaitoni L 2014 Minimum energy of computing, fundamental considerations *ICT—Energy—Concepts Towards Zero-Power Information and Communication Technology* (Rijeka: InTech) ch 7
[10] Borkar S 2006 Electronics Beyond Nano-scale CMOS Design Automation Conference 43rd ACM/IEEE pp 807–8
[11] Gardiner C W 1985 *Handbook of Stochastic Methods* (Berlin: Springer)
[12] Marconi U M B, Puglisi A, Rondoni L and Vulpiani A 2008 *Phys. Rep.* **461** 111195
[13] Berut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R and Lutz E 2012 *Nature* **483** 187–189
[14] Lambson B, Carlton D and Bokor J 2011 *Phys. Rev. Lett.* **107** 010604
[15] Madami M, Gubbiotti G, Tacchi S and Carlotti G 2013 *Nanoenergy Lett* **6** 27
[16] von Neumann J 1966 *Fourth University of Illinois lecture, in Theory of Self-Reproducing Automata* ed A W Burks (Urbana, IL: University of Illinois Press) p 66
[17] Landauer R 1961 *IBM J. Res. Dev.* **5** 183–91
[18] Bennett C H 1982 *Int. J. Theor. Phys.* **21** 905–40
[19] Maroney O J E 2009 *Phys. Rev. E* **031105
[20] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
[21] Orlov A O, Lent C S, Thorpe C C, Boechler G P and Snider G L 2012 *Japan. J. Appl. Phys.* **51** 06FE10
[22] Toyabe S, Sagawa T, Ueda M, Muneyuki E and Sano M 2010 *Nat. Phys.* **6** 988992
[23] Gammaitoni L 2007 *Appl. Phys. Lett.* **91** 224104
[24] Gammaitoni L 2011 arXiv:1111.2937
[25] Gammaitoni L 2013 *Nanoenergy Lett.* **5** 10
[25] Gammaitoni L, Hanggi P, Jung P and Marchesoni F 1998 *Rev. Mod. Phys.* **70** 223