Fundamental energy dissipation limits in ICT devices

Luca Gammaitoni
NiPS Laboratory, University of Perugia (IT)

With:
Davide Chiucchiù (now at Okinawa uni.), Miquel Lopez (now at ICMAB), Cristina Diamantini and Igor Neri at NiPS.

Aalto University, May 25th 2023

NiPS Laboratory
Noise in Physical Systems

University of Perugia (IT)

AD 1308

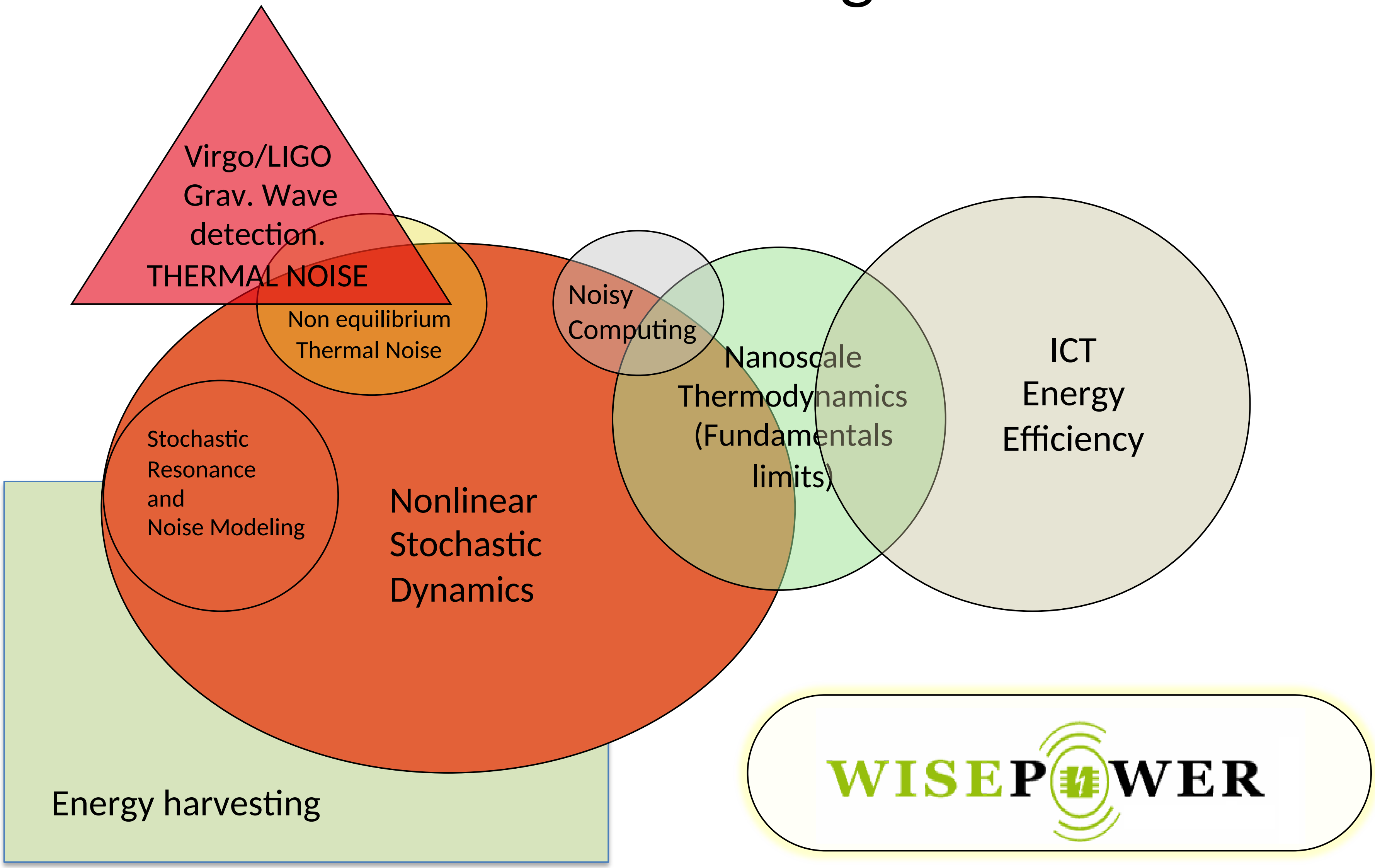
NiPS Laboratory
Noise in Physical Systems

Cristina Diamantini, Francesco Cottone,
Igor Neri, Alessandro di Michele, Maurizio Mattarelli,
Giacomo Clementi, Raffella Pellegrini, Luca
Gammaitoni, Paolina Cerlini

ICT-Energy ZEROPOWER

www.nipslab.org

What are we doing at NiPS?



NiPS Laboratory
Noise in Physical Systems

We are interested in noise and fluctuations.
Energy transformation processes at micro and nano scales.

ICT-Energy
Fundamental limits in the physics of computing

Questions like:

- Can we operate a computer by spending 0 energy?
- How long can a memory last?
- How much energy dos it take to remember something?

Funfed projects

2006-2009 EC (SUBTLE VIFP)
2010-2013 EC (NANOPOWER VIIFP)
2010-2013 EC (ZEROPOWER VIIFP)
2012-2015 EC (LANDAUER VIIFP)
2013-2016 EC (ICT-Energy VIIFP)
2015-2018 EC (Proteus H2020)
2017-2020 EC (OPRECOMP H2020)
2017-2021 EC (ENABLES H2020)
2022-2026 PNRR VITALITY (NextGenerationEU)

www.nipslab.org

ICT-Energy ZEROPOWER

NiPS Laboratory
Noise in Physical Systems

April 4, 2005

March 13, 2013

Our agenda

1 Computers are energy hungry

2 Basic principles in computing

3 Fundamental limits in energy consumption

Our agenda

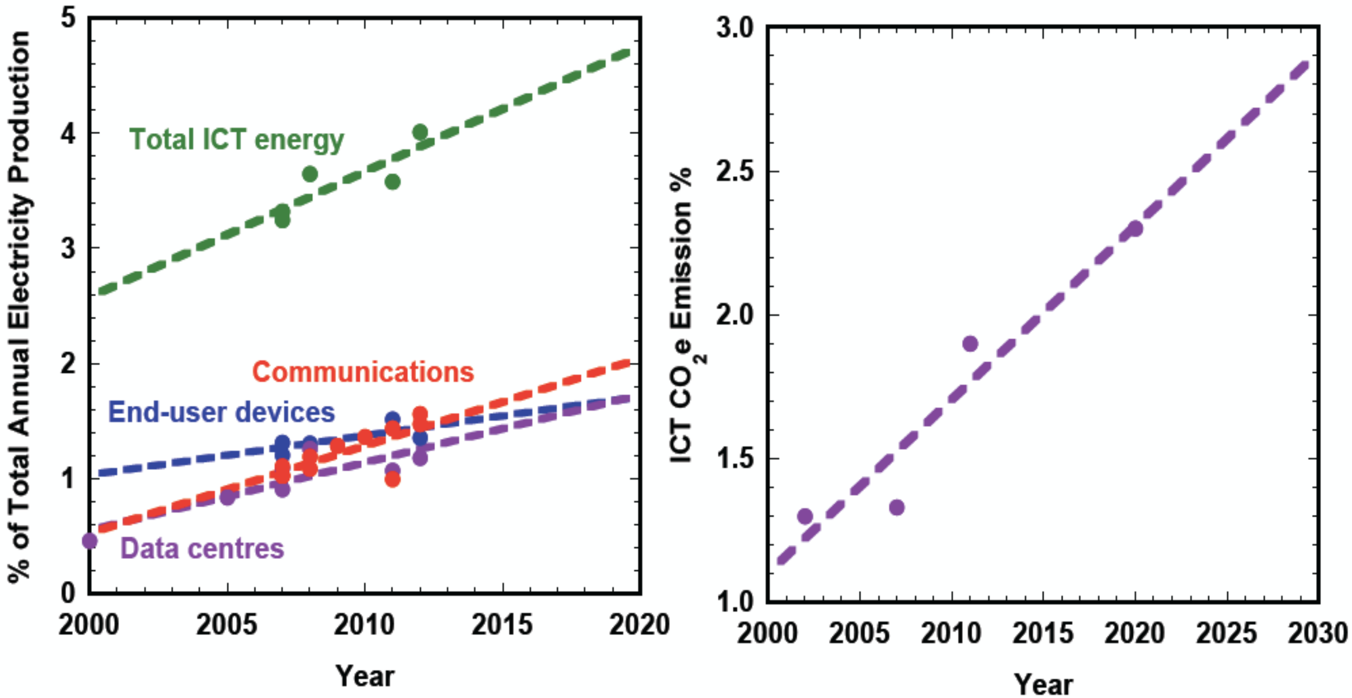
1 Computers are energy hungry

2 Basic principles in computing

3 Fundamental limits in energy consumption

1 Computers are energy hungry

ICT global energy consumption



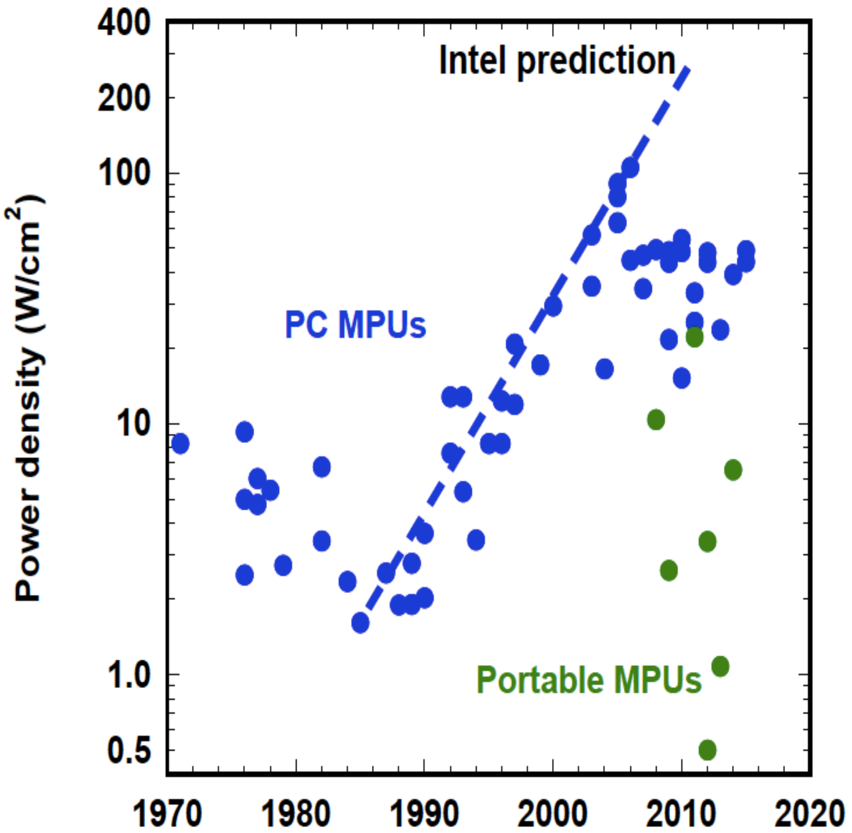
- **This excludes TV, media, publishing, games, power switches, domestic & industrial ICT devices**

Source: D. Paul, ICT-Energy Research Agenda, 2015

NiPS Laboratory
Noise in Physical Systems

1 Computers are energy hungry

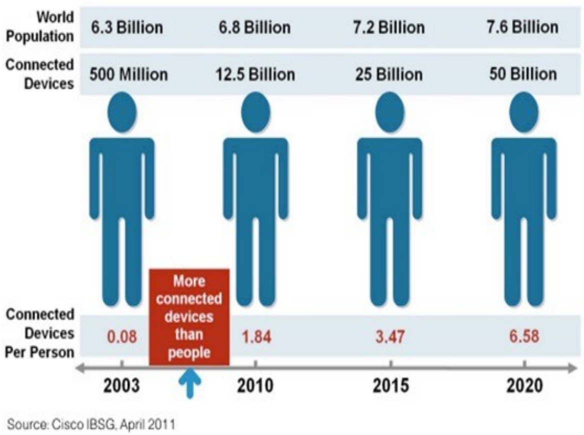
If we want more powerful supercomputers reducing energy is strategic



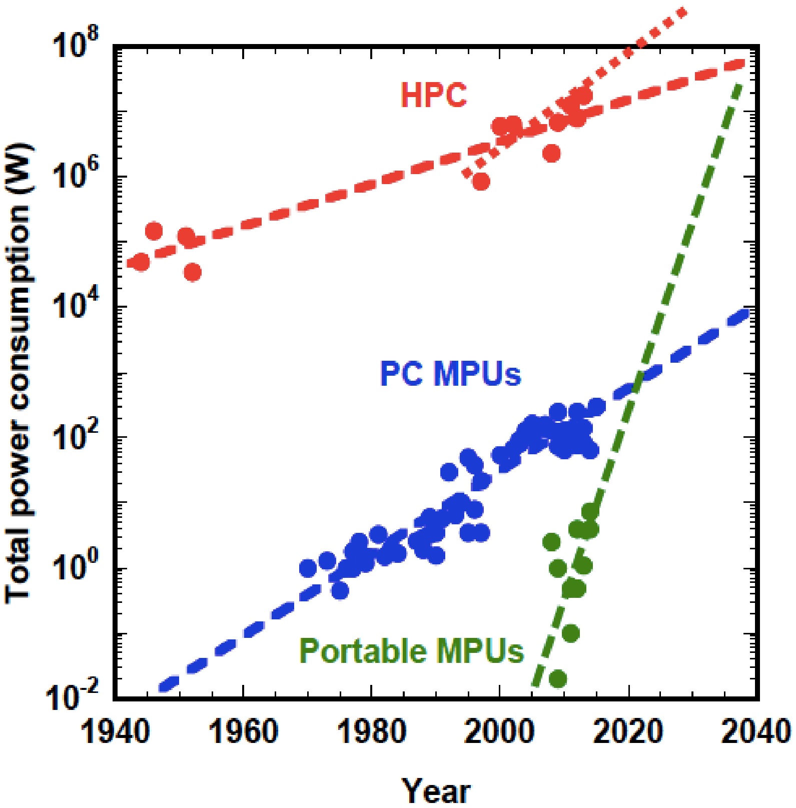
Source: D. Paul, ICT-Energy Research Agenda, 2015

1 Computers are energy hungry

If we want the **Internet of Things** to happen



...to avoid lacking autonomous power



Source: D. Paul, ICT-Energy Research Agenda, 2015

Our agenda

1 Computers are energy hungry

2 Basic principles in computing

3 Fundamental limits in energy consumption

Computing devices as physical systems

In order to be able to produce an observable change in the system, the system must admit at least 2 different states.

A computation is associated with system transformation

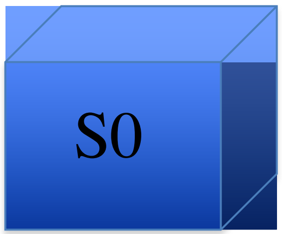
System transformations are described by Physics

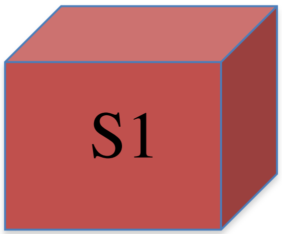
A physical system that can assume two distinguishable states is called a **binary switch**

Let’s start with some basic modeling

One dimensional dynamical system x(t)

x(t) is a physical observable and we can identify two states.





S0 if $x < x_{TH}$

S1 if $x > x_{TH}$

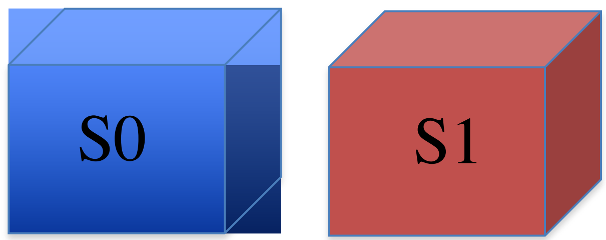
We need a “change rule”:

Newton equation approach

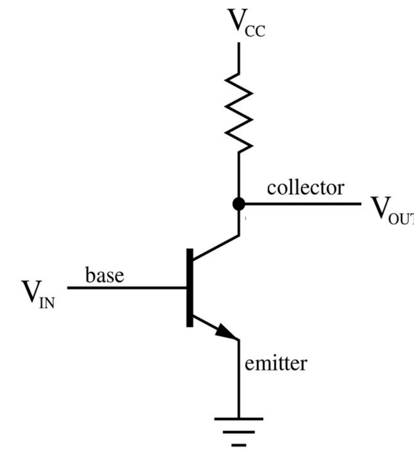
$$x = -\frac{dU}{dx} + F_{!"\#}$$

Where U(x) is a confining potential and F_{ext} is the force that can make the state change possible.

A simple electronic model



S0 if $V_{OUT} < V_T$ S1 if $V_{OUT} > V_T$



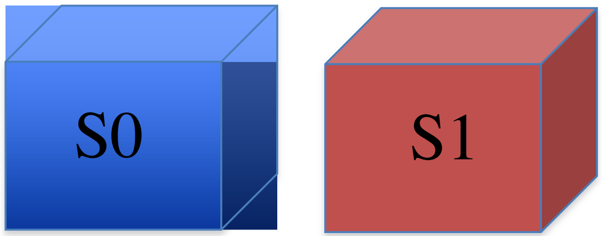
One dimensional dynamical system x(t)

x(t) represents the electric voltage at a given point

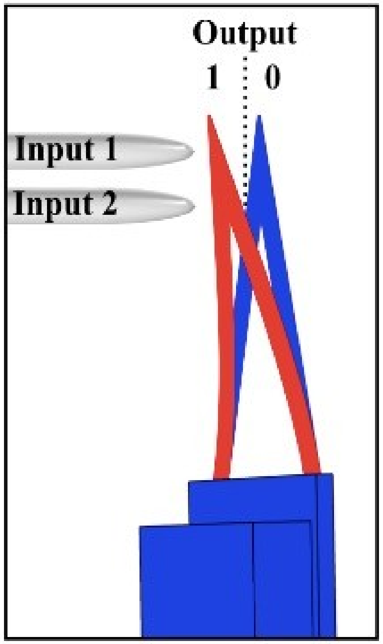
The two states are represented by the measurable quantity “electric voltage” at point V_{OUT}. As an example state “S0” = $V_{OUT} < V_T$; state “S1” = $V_{OUT} > V_T$; with V_T a given reference voltage.

The way to induce state changes represented by an electromotive force applied at point V_{IN}.

A simple mechanical model



S0 if $x < x_{TH}$ S1 if $x > x_{Th}$



One dimensional dynamical system x(t)

x(t) represents the position of the tip of a microcantilever

The two states are represented by the measurable quantity "position of the tip". As an example state "S0" = $x < x_{TH}$; state "S1" = $x > x_{TH}$; with x_{Th} a given reference position.

The way to induce state changes represented by an electrostatic force applied from outside.

In general we have two classes of binary switches: ***combinational*** and ***sequential***

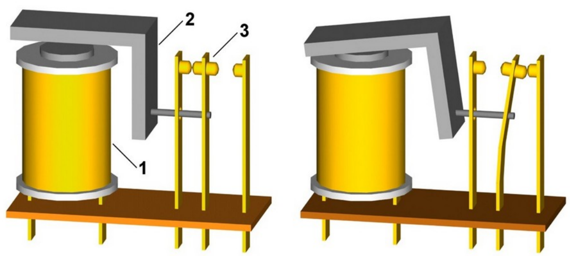
Conbinational:

in the absence of any external force, under equilibrium conditions, they are in the state S_0. When an external force F_{01} is applied, 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 go back to the state S_0.

Sequential:

They can be changed from S_0 to S_1 by applying an external force F_{01}. Once they are in the state S_1 they remain in this state even when the force is removed. They go from S_1 to S_0 by applying a new force F_{10}. Once they are in S_0 they remain in this state even when the force is removed.

Example



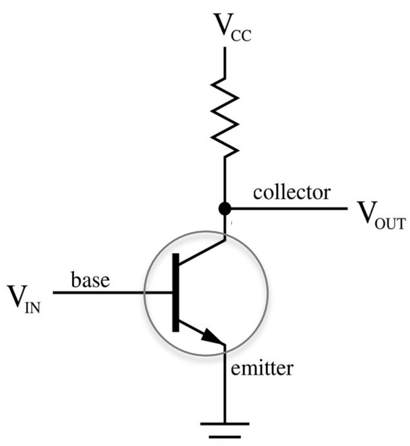
Example

The previous dynamical model is useful for both classes of binary switches

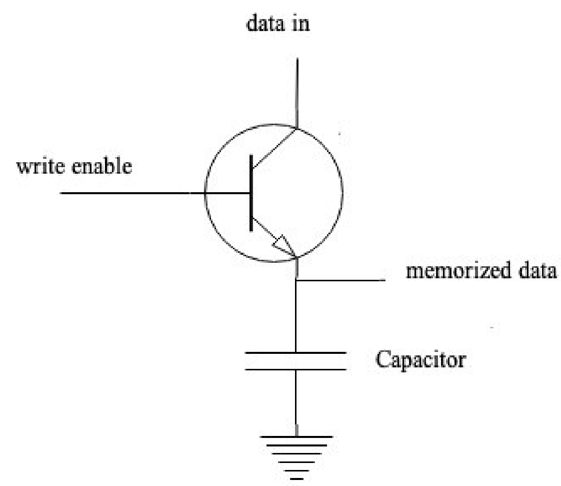
In modern computers binary switches are made with transistors. These are electronic devices that satisfy the two conditions required:

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.

The way to induce state changes represented by an electromotive force applied at point V_{IN}.



Combinational switch

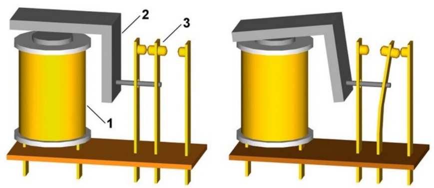


Sequential switch

Minimum Energy of Computing, Fundamental Considerations, Victor Zhirnov, Ralph Cavin and Luca Gammaitoni
in the book "ICT - Energy - Concepts Towards Zero - Power Information and Communication Technology”, InTech, 2014

Both classes can be described by the same equation $x = -\frac{dU}{dx} + F_{!"\#}$

How do we distinguish them in our model ?

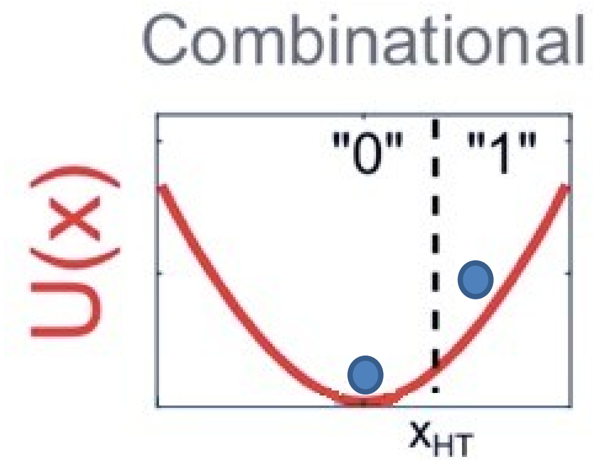


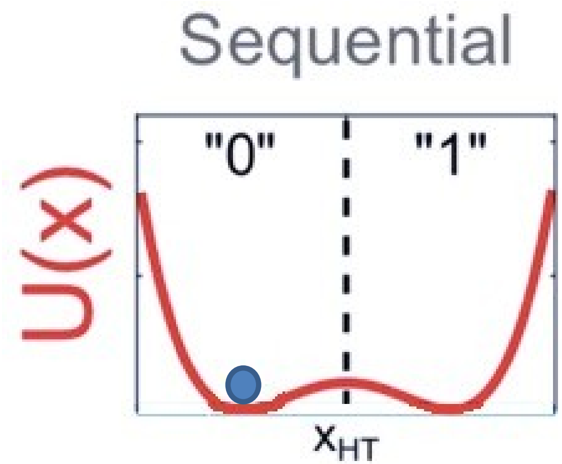
Combinational

Sequential

$$U(x) = a\frac{1}{2}x^2$$

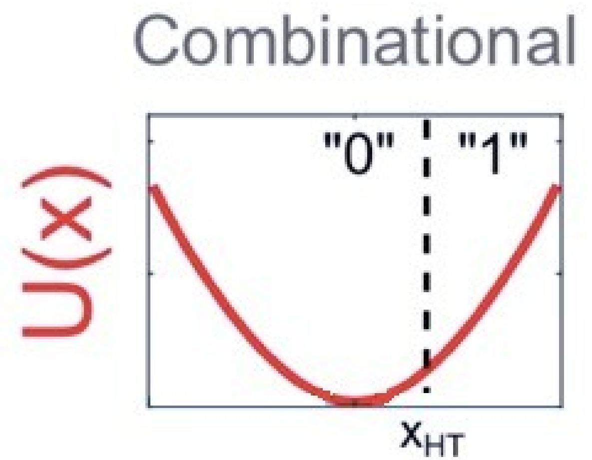
$$U(x) = -a\frac{1}{2}x^2 + b\frac{1}{4}x^4$$

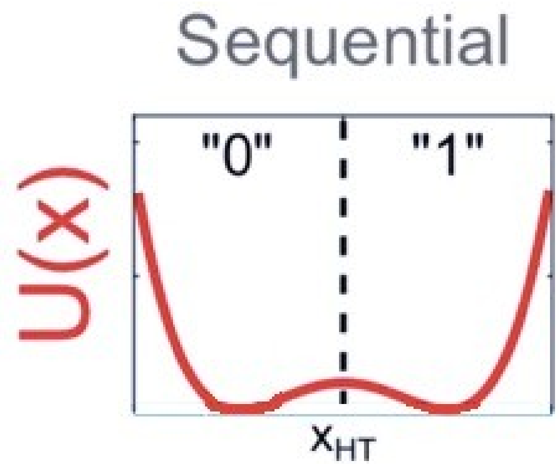




So, this is our dynamical model for the binary switches $$x = - \frac{dU}{dx} + F_{!"\#}$$

$$U(x) = a\frac{1}{2}x^2 \qquad U(x) = -a\frac{1}{2}x^2 + b\frac{1}{4}x^4$$





Is this equation enough to describe the dynamics of our switch?

Physical systems whose dynamical behavior can be described in the framework of non-equilibrium statistical mechanics.

Langevin equation approach $m\ddot{x} = -\gamma\dot{x} + \zeta + F_{ext}$

$$F_{ext} = -\frac{dU(x,t)}{dx} + \zeta$$

Deterministic force depending on x, t

Random force depending on t

This last equation is more complicated that the previous one but is more realistic.

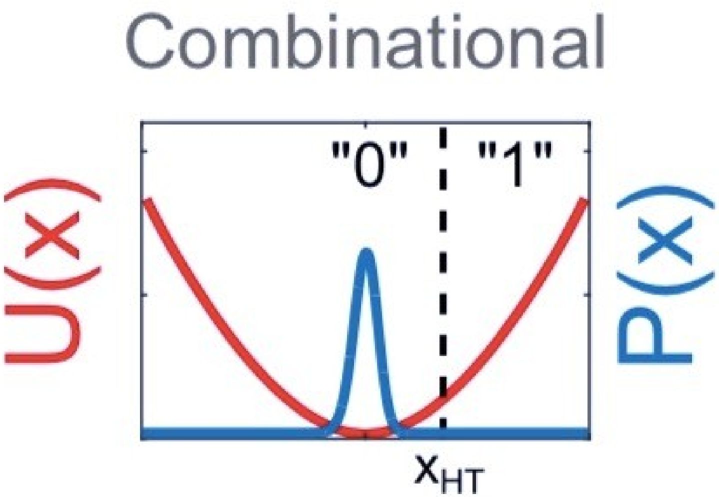
Question: **how do we describe now the behaviour of x(t) ?**

Back to our model sw

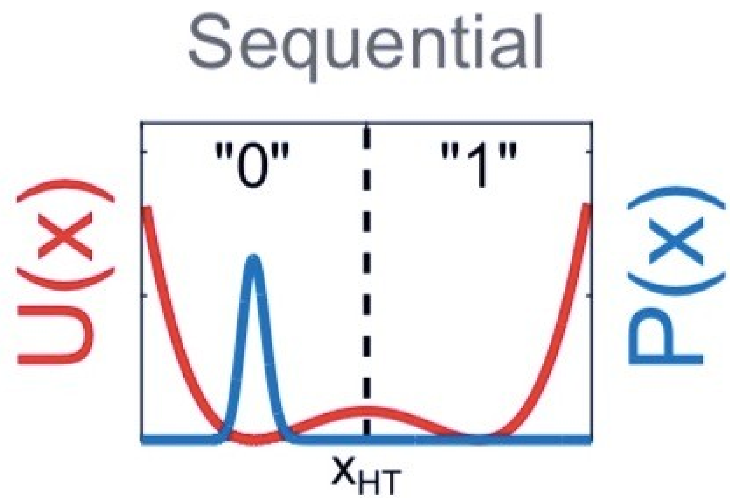
$$\ddot{x} = -\frac{dU(x)}{dx} - \gamma\dot{x} + F(t) + \sigma\xi(t)$$

Due to the presence of the fluctuations we need to introduce the probability density P(x)

$$U(x) = a\frac{1}{2}x^2$$



$$U(x) = -a\frac{1}{2}x^2 + b\frac{1}{4}x^4$$



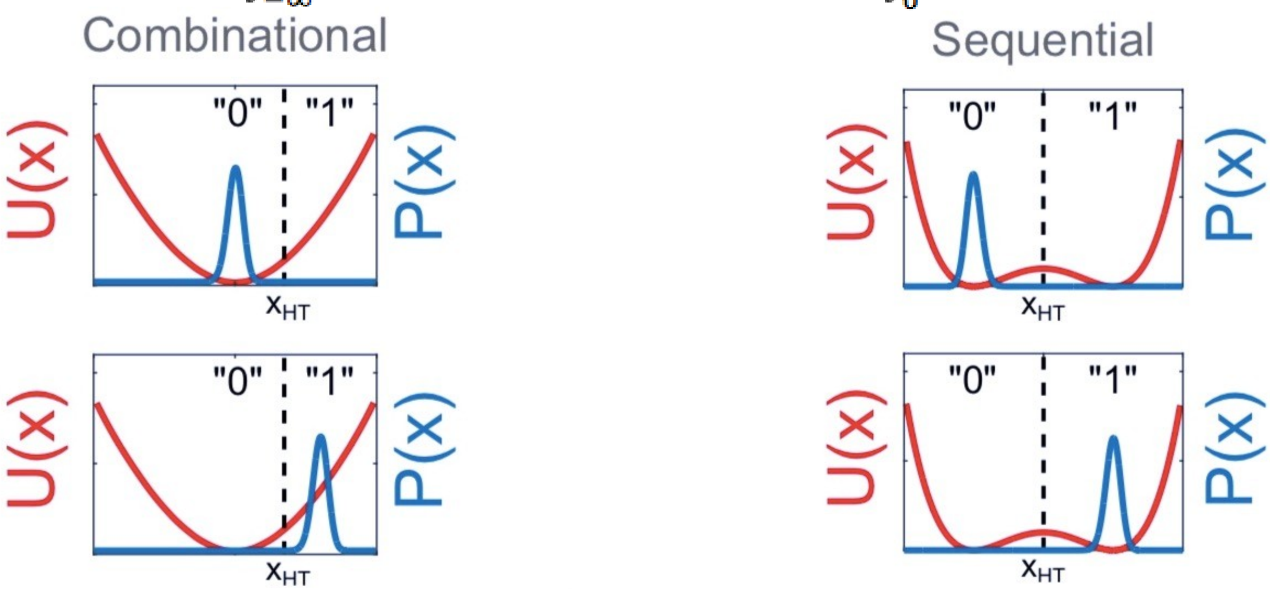
The Physics of switches: the switch event

Based on these considerations we now define the switch event as the transition from an initial condition toward a final condition, where the initial condition is defined as <x> < 0 and the final condition is defined as <x> > 0. With the initial condition characterized by:

$$p_0(t) = \int_{-\infty}^{0} \boldsymbol{P}(\boldsymbol{x},\boldsymbol{t})dx \cong 1 \quad and \quad p_1(t) = \int_{0}^{+\infty} \boldsymbol{P}(\boldsymbol{x},\boldsymbol{t})dx \cong 0$$

and the final condition by:

$$p_0(t) = \int_{-\infty}^{0} \boldsymbol{P}(\boldsymbol{x},\boldsymbol{t})dx \cong 0 \quad and \quad p_1(t) = \int_{0}^{+\infty} \boldsymbol{P}(\boldsymbol{x},\boldsymbol{t})dx \cong 1$$



Our agenda

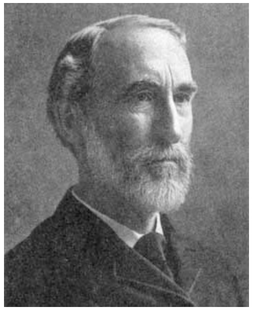
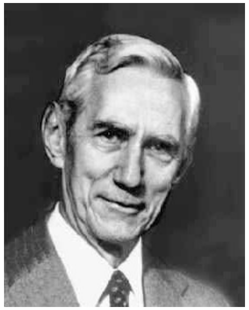
1 Computers are energy hungry

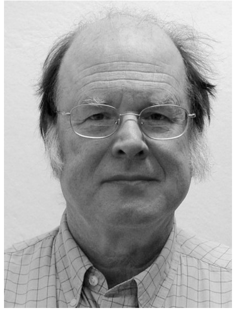
2 Basic principles in computing

3 Fundamental limits in energy consumption

Questions

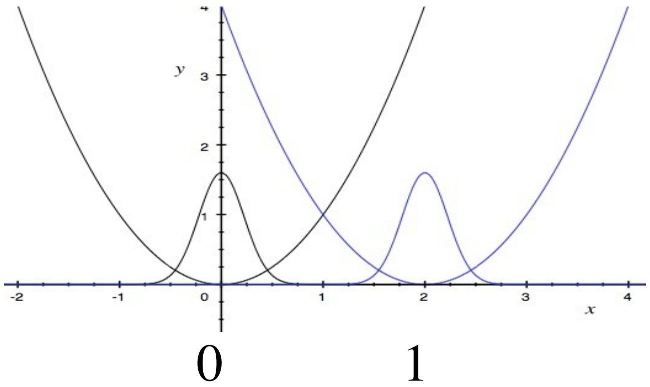
- What is the minimum energy we have to spend if we want to produce a switch event ?
- Does this energy depends on the technology of the switch ?
- Does this energy depends on the instruction that we give to the switch ?
- ….

There is one basic operation we can do with a **combinational switch**

The **switch operation** (i.e. the change of state)

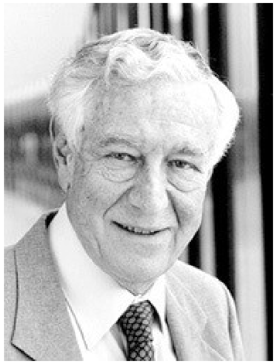


Before the switch = 1 logic state
After the switch = 1 logic state

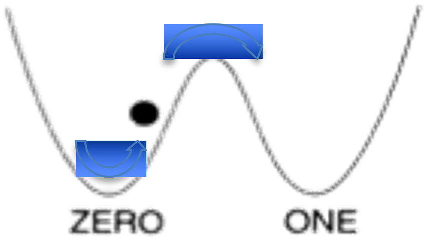
Change in entropy = $S_f - S_i = K_B \log(1) - K_B \log(1) = 0$

No net decrease in entropy ---> no minimum energy required

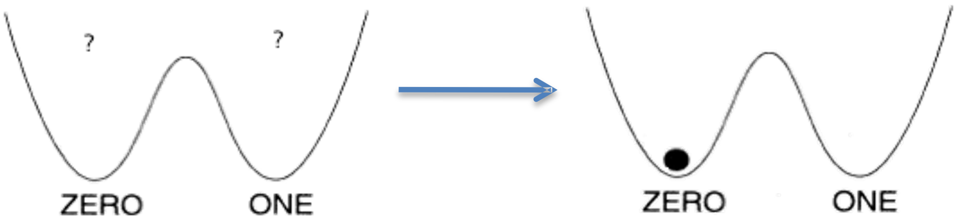
There are two basic operations we can do with a **sequential switch**



The **switch operation** (i.e. the change of state)

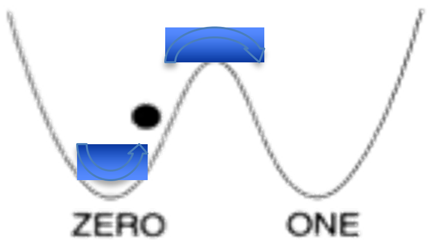


The **reset operation** (i.e. the set of a given state starting from an unknown state)



Let’s look at this, with a reasoning introduced in 1961 by R. Landauer

The single switch operation



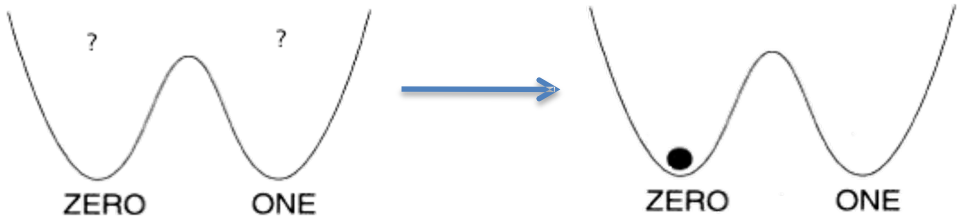
$0 \longrightarrow 1$

Before the switch = 1 logic state
After the switch = 1 logic state

Change in entropy = $S_f - S_i = K_B \log(1) - K_B \log(1) = 0$

No net decrease in entropy ---> no minimum energy required

The reset operation



? → 0

Before the reset = 2 possible logic states
After the reset = 1 logic state

Change in entropy = $S_f - S_i = K_B \log(1) - K_B \log(2) = - K_B \log(2)$

Net decrease in entropy ---> **energy expenditure required**

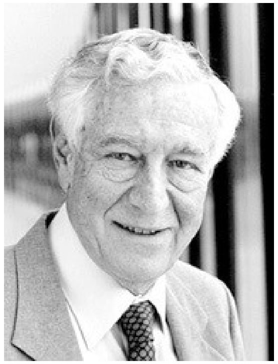
THE VON NEUMANN-LANDAUER BOUND

The Landauer's principle (1) states that erasing one bit of information (like in a resetting operation) comes unavoidably with a decrease in physical entropy and thus is accompanied by a minimal dissipation of energy equal to

$$Q = k_B\, T \ln 2$$

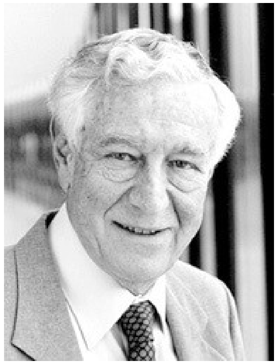
More technically this is the result of a change in entropy due to a change from a random state to a defined state.

Please note: this is the **minimum** energy required.



(1) R. Landauer, "Dissipation and Heat Generation in the Computing Process" IBM J. Research and Develop. 5, 183-191 (1961),

LANDAUER'S TAKE

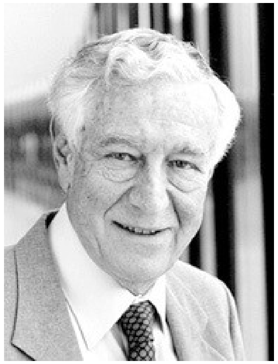


In the same paper Landauer generalized this result associated with the reset operation to the cases where there was a decrease of information between the input and the output of a computing system.
Landauer wrote (1):

We shall call a device logically irreversible if the output of a device does not uniquely define the inputs. We believe that devices exhibiting logical irreversibility are essential to computing. Logical irreversibility, we believe, in turn implies physical irreversibility, and the latter is accompanied by dissipative effects.

Three sentences: one definition and two beliefs.

(1) R. Landauer, "Dissipation and Heat Generation in the Computing Process"
IBM J. Research and Develop. 5, 183-191 (1961),



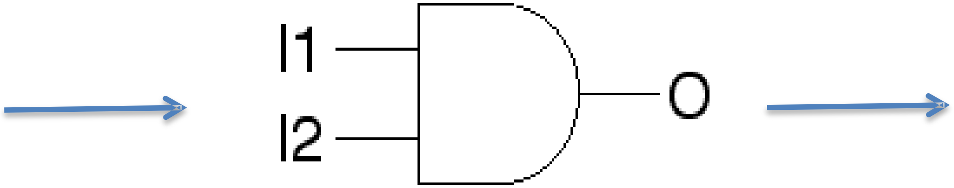
The definition: Logical irreversibility

We shall call a device logically irreversible if the output of a device does not uniquely define the inputs. We believe that devices exhibiting logical irreversibility are essential to computing. Logical irreversibility, we believe, in turn implies physical irreversibility, and the latter is accompanied by dissipative effects.

I_{in} Information in

I_{out} Information out

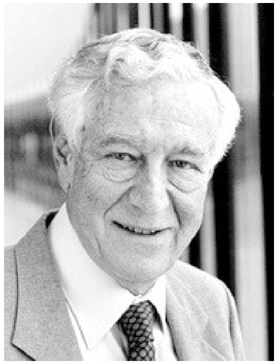
Logically irreversible $I_{out} < I_{in}$



I_{in} Information in

I_{out} Information out

The first belief: Logical irreversibility is necessary



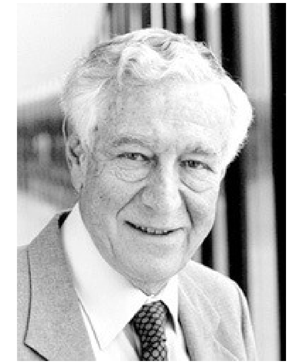
We shall call a device logically irreversible if the output of a device does not uniquely define the inputs. We believe that devices exhibiting logical irreversibility are essential to computing. Logical irreversibility, we believe, in turn implies physical irreversibility, and the latter is accompanied by dissipative effects.

Logically reversible COMPUTING ($I_{out} = I_{in}$)

Can be used to do computation

C. H. Bennett, "Logical reversibility of computation," IBM Journal of Research and Development, vol. 17, no. 6, pp. 525-532, 1973.

The second belief: Logical irreversibility -> dissipation



We shall call a device logically irreversible if the output of a device does not uniquely define the inputs. We believe that devices exhibiting logical irreversibility are essential to computing. Logical irreversibility, we believe, in turn implies physical irreversibility, and the latter is accompanied by dissipative effects.

Landauer paradigm

I_{in} Information in → Computing device → I_{out} Information out

Thermodynamics paradigm

Computing device → (time) → Computing device

S_i initial entropy

S_f final entropy

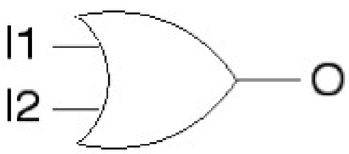
Testing the logical irreversibility limit

I_{in} Information in

I_{out} Information out

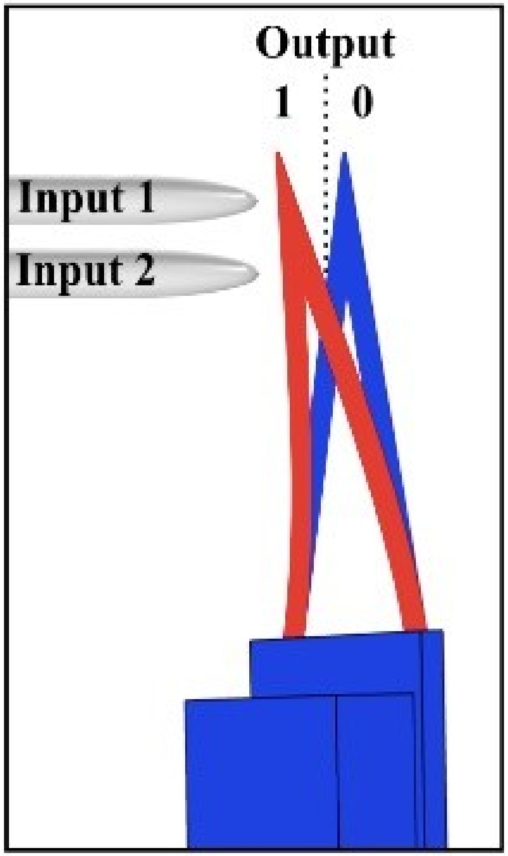
Logical irreversibility: $I_{out} < I_{in}$

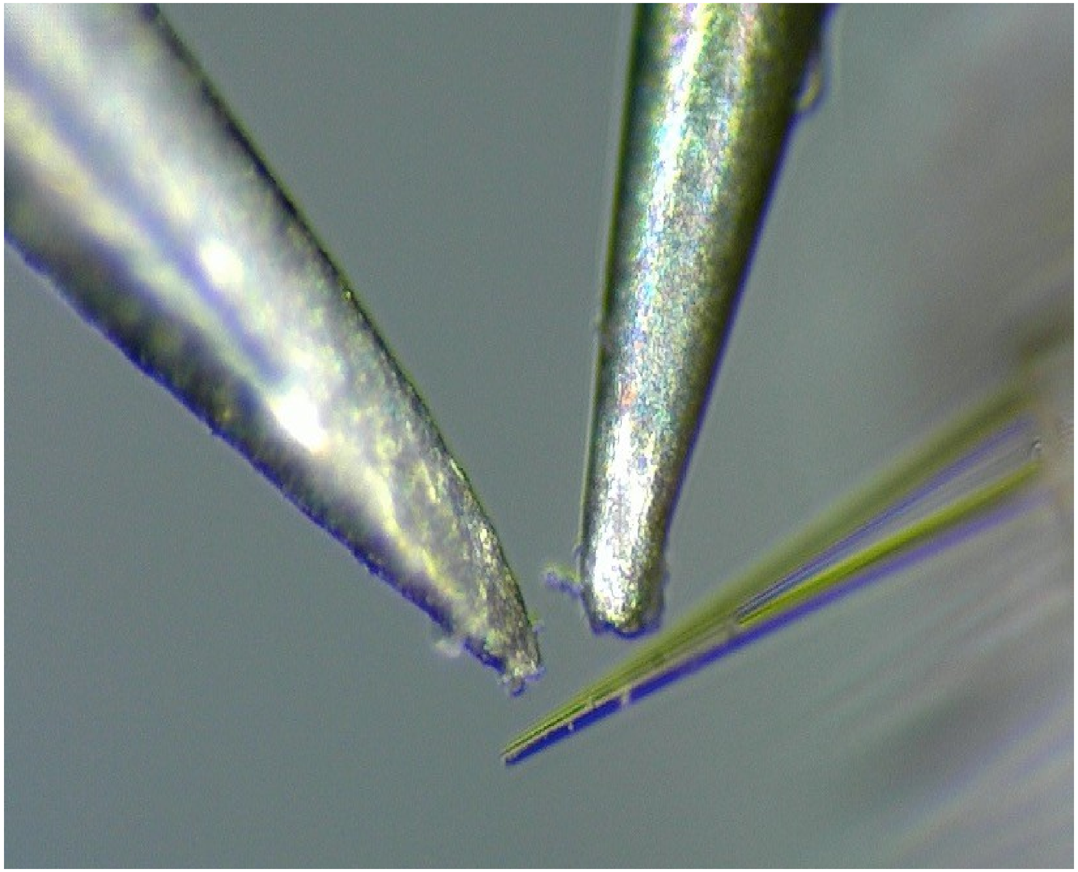
OR



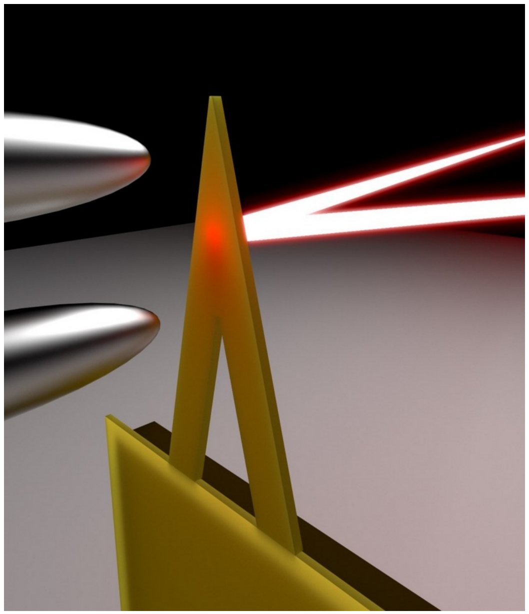
I1	I2	O
0	0	0
0	1	1
1	0	1
1	1	1

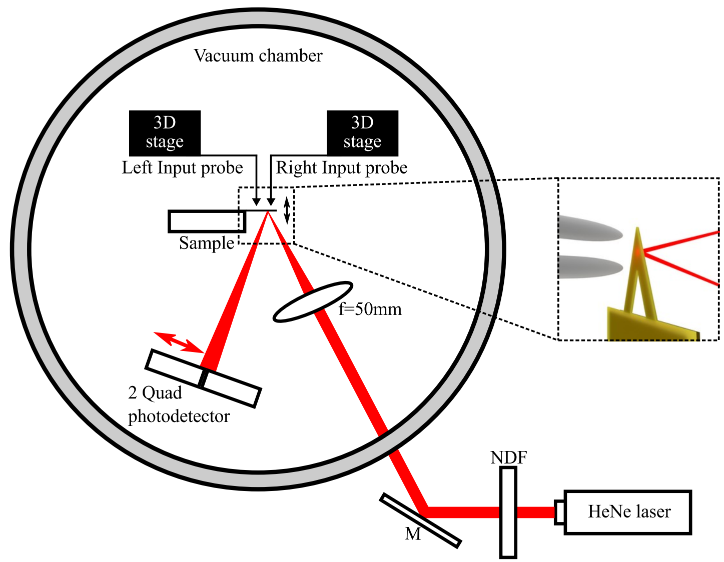
Micro electro-mechanical Logic gate





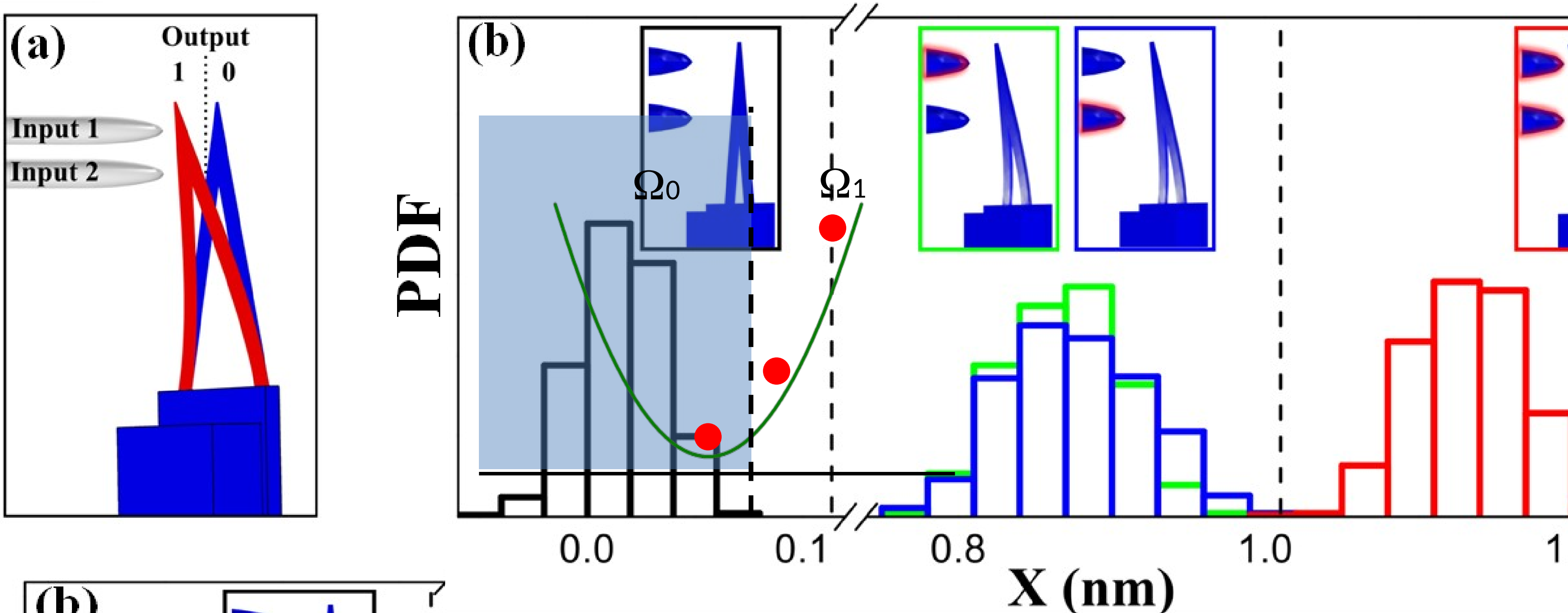
Micro electro-mechanical Logic gate

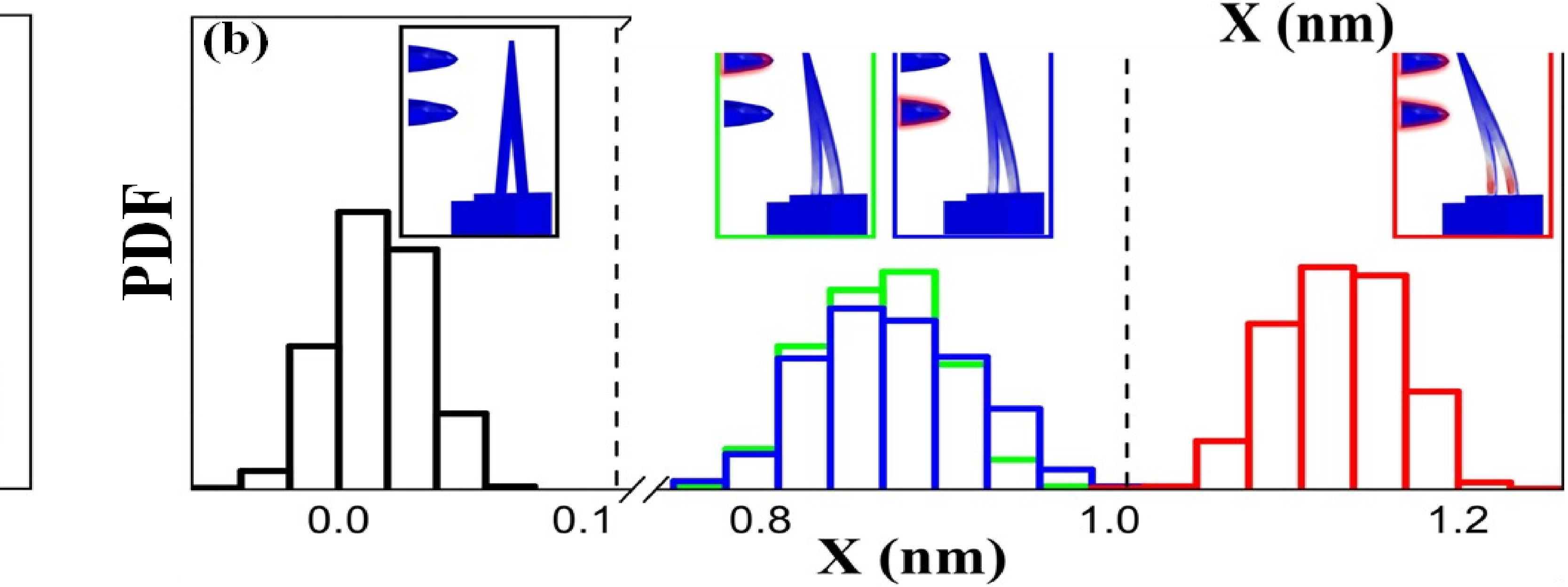




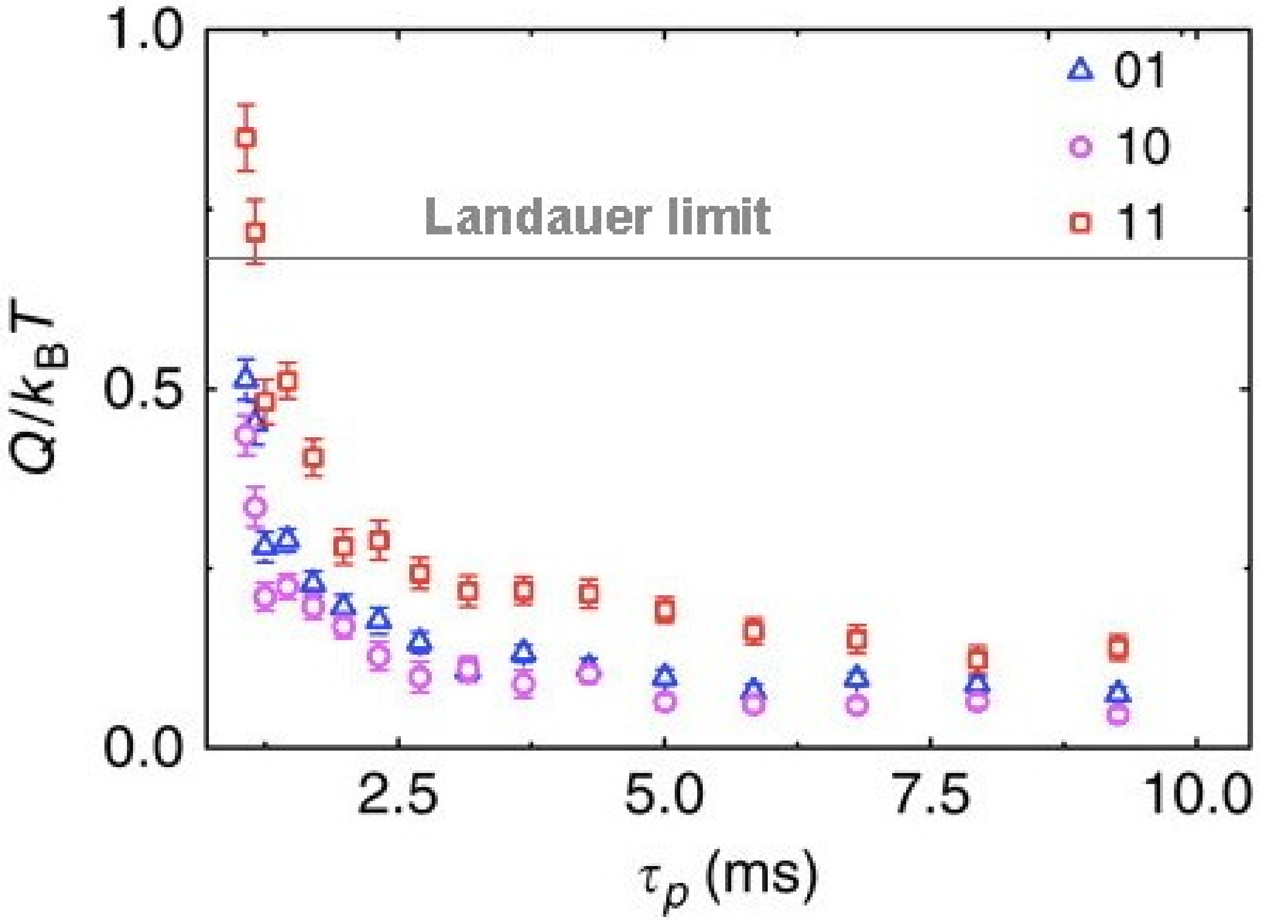
All measurements are carried on in a vacuum chamber at P = 10^{-3}±0.01 mbar and at room temperature (T = 300 K). The mechanical structure is a 200 ∝m long V-shaped cantilever, with a first-mode resonant frequency of fr = 14,950±1 Hz and a quality factor Q = 2,886±10, resulting in a relaxation time t = 61.4±0.2 ms.

The deflection of the cantilever, x, is measured by an atomic force microscopy-like optical lever: the deflection of the laser beam (633 nm) due to the bend of the cantilever is detected by a two quadrants photo detector. Position and voltage measurements were digitalized at 50 kHz.


(a)
Output
1
0
Input 1
Input 2
(b)
PDF
Ω_0
Ω_1
0.0
0.1
0.8
1.0
1
X (nm)

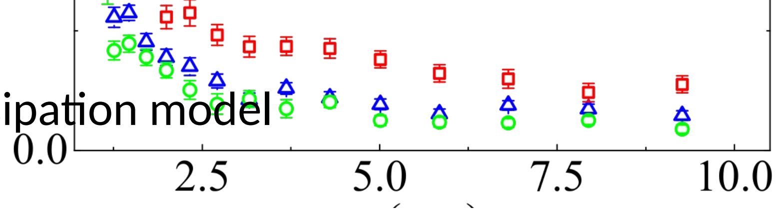

(b)
PDF
0.0
0.1
0.8
1.0
1.2
X (nm)

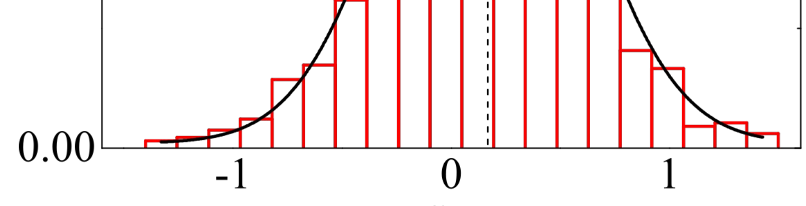
Measure of the energy dissipated during information processing with OR logic gate

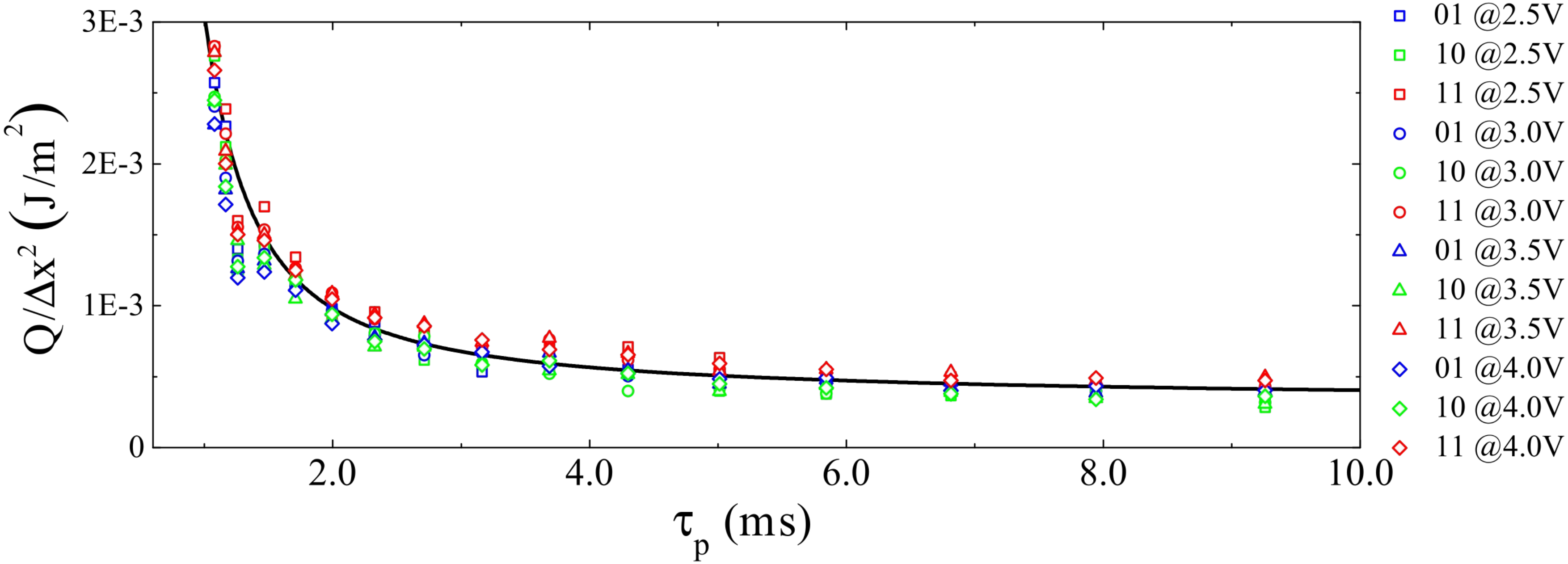


Sub-kBT micro-electromechanical irreversible logic gate,
M. López-Suárez, I. Neri, L. Gammaitoni. Nature Communications 7, Article number: 12068 (2016)

Dissipation model



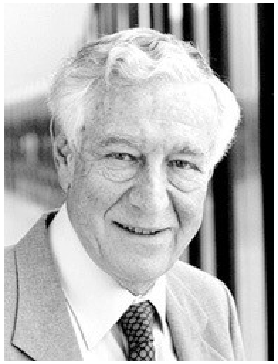




Zener theory $-k(1+i\phi)$

$$\phi(\nu) = \phi_{\mathrm{str}} + \phi_{\mathrm{th-el}} + \phi_{\mathrm{vis}} + \phi_{\mathrm{clamp}}$$

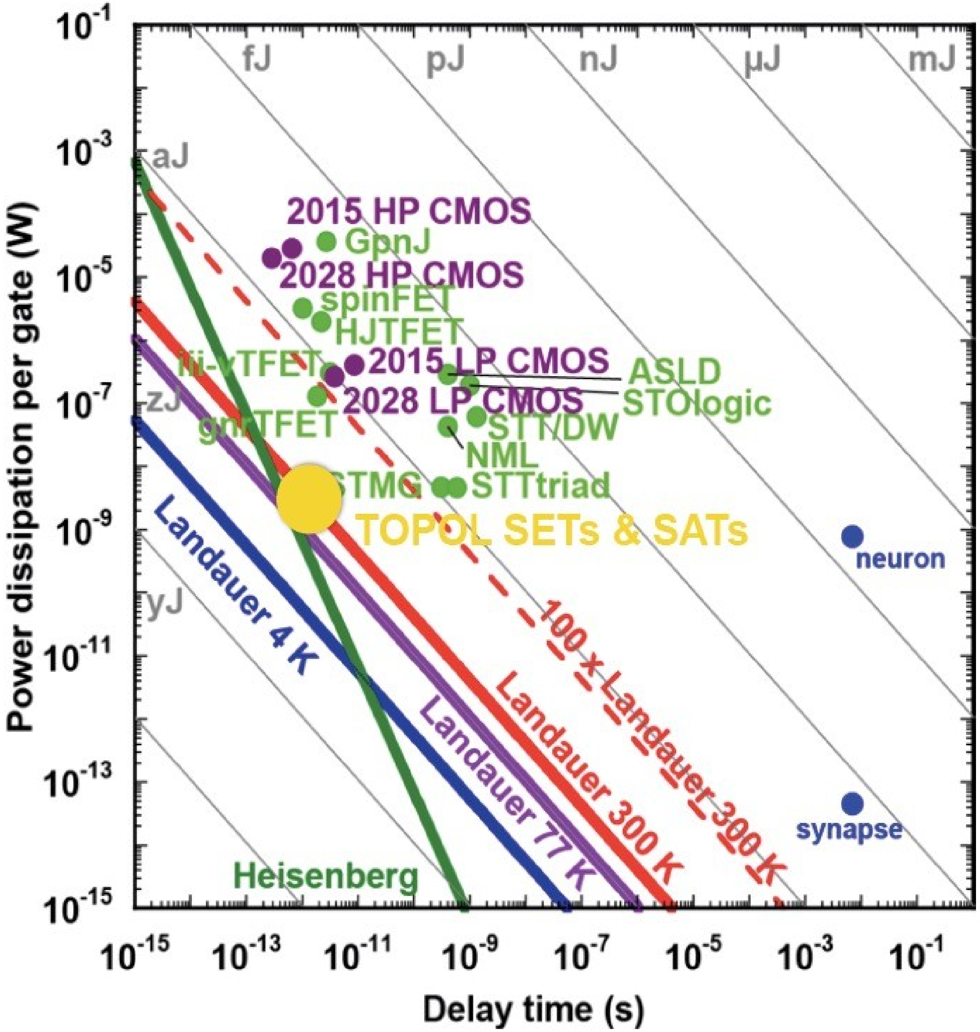
The second belief: Logical irreversibility -> dissipation



We shall call a device logically irreversible if the output of a device does not uniquely define the inputs. We believe that devices exhibiting logical irreversibility are essential to computing. **Logical irreversibility, we believe, in turn implies physical irreversibility, and the latter is accompanied by dissipative effects**.

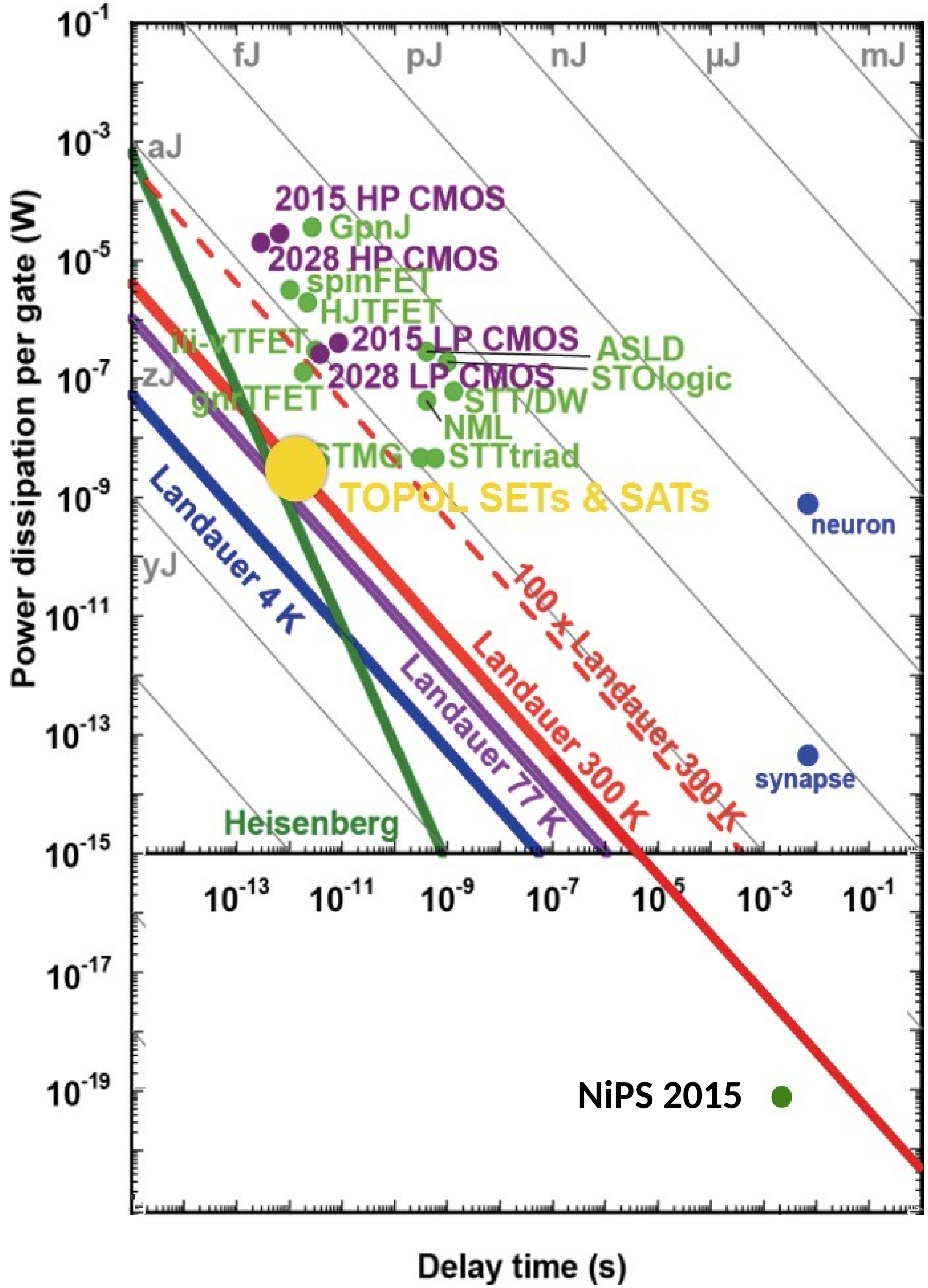
This is not apparently the case.
Logical reversibility is not needed in order to perform zero-dissipation computing.

The state of the art in energy dissipation during computation



Source: D. Paul, ICT-Energy Research Agenda, 2015

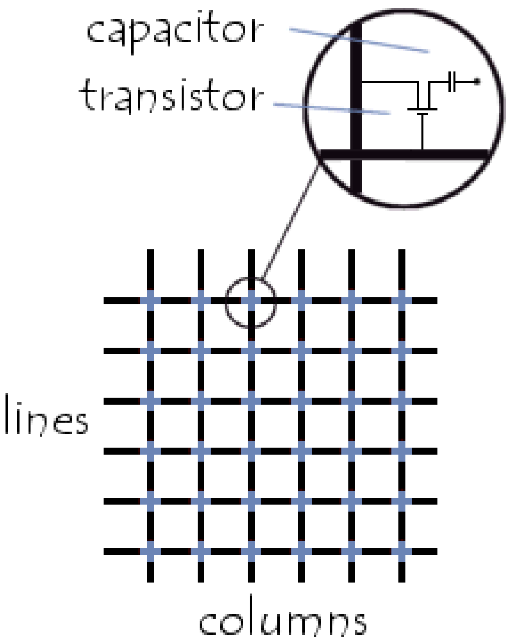
The state of the art in energy dissipation during computation

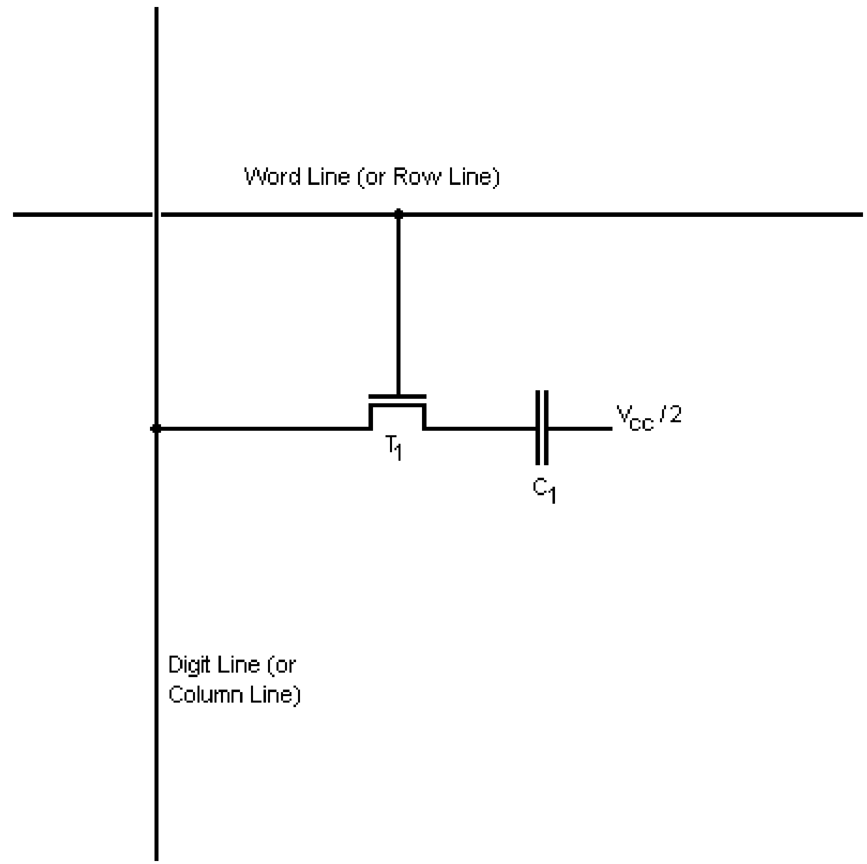


Noise in Physical Systems

Source: D. Paul, ICT-Energy Research Agenda, 2015

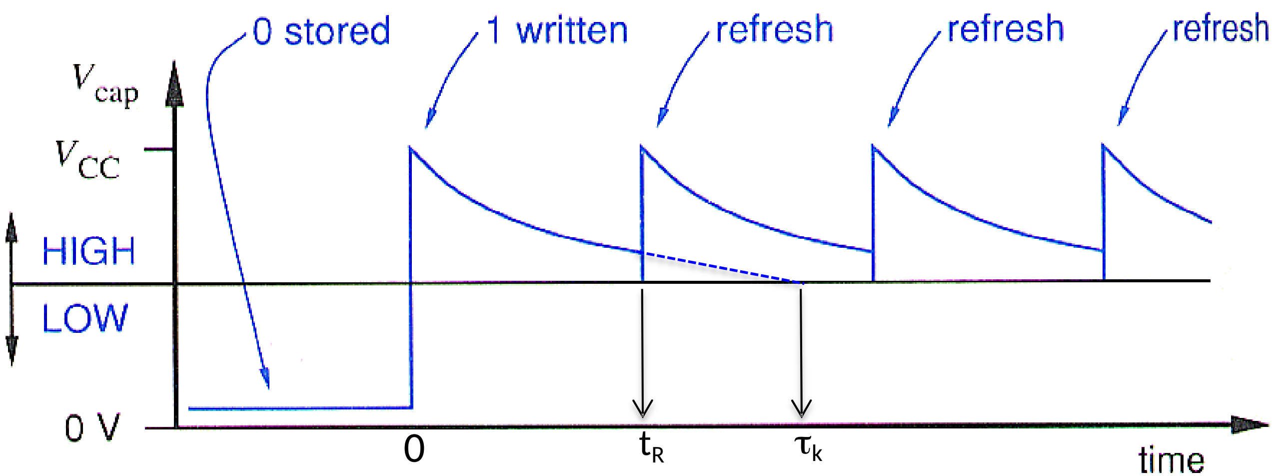
What about computer memories ?





DRAM

In order to counterbalance the memory degradation, a periodic refresh operation is performed



If no refresh operation is performed the memory is lost on average after a time τ_K

The refresh operation is performed periodically with period t_R

The refresh operation last for a time t_p

$t_p << t_R << \tau_K$

Nano seconds → t_p

Micro seconds → t_R

seconds → τ_K

Scope of the work

Assumed that the refresh operation has an energetic cost $\mathbf{Q}$

we are interested in the **fundamental energy limits** to preserve a given bit for a time $\mathbf{t}$

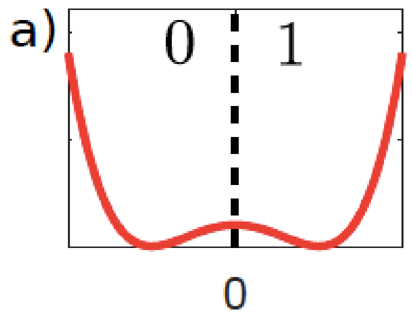
with a probability of failure not larger than $\mathbf{P_E}$

while executing the refresh procedure with periodicity $\mathbf{t_R}$

Plan of the work

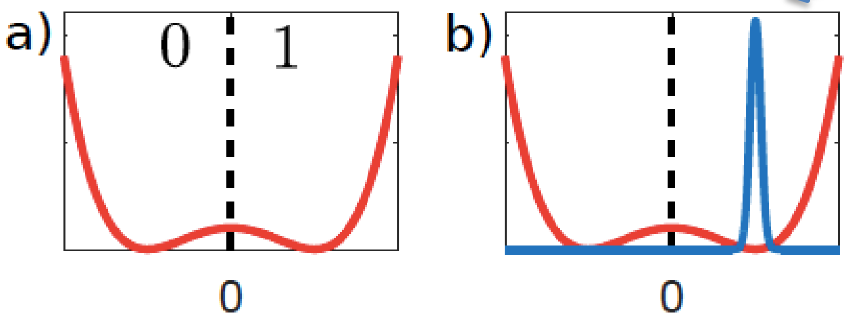
1. introduce a simple physical model for the 1-bit memory
2. Compute $\mathbf{t_R}$ for a given set of $\mathbf{P_E}$ and $\mathbf{t}$
3. Perform an experiment to determine the minimum energy required
4. Elaborate considerations

1 1-bit memory



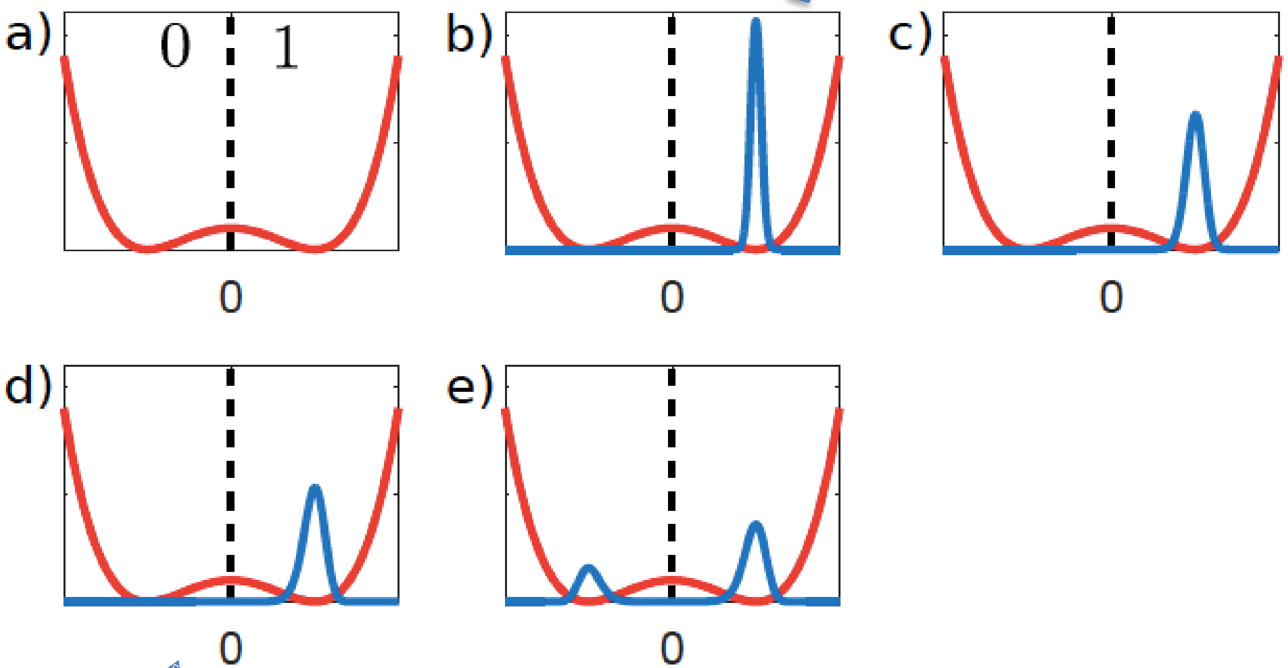
1 1-bit memory

Initial probability density p(x,t)



1 1-bit memory

Initial probability density p(x,t)

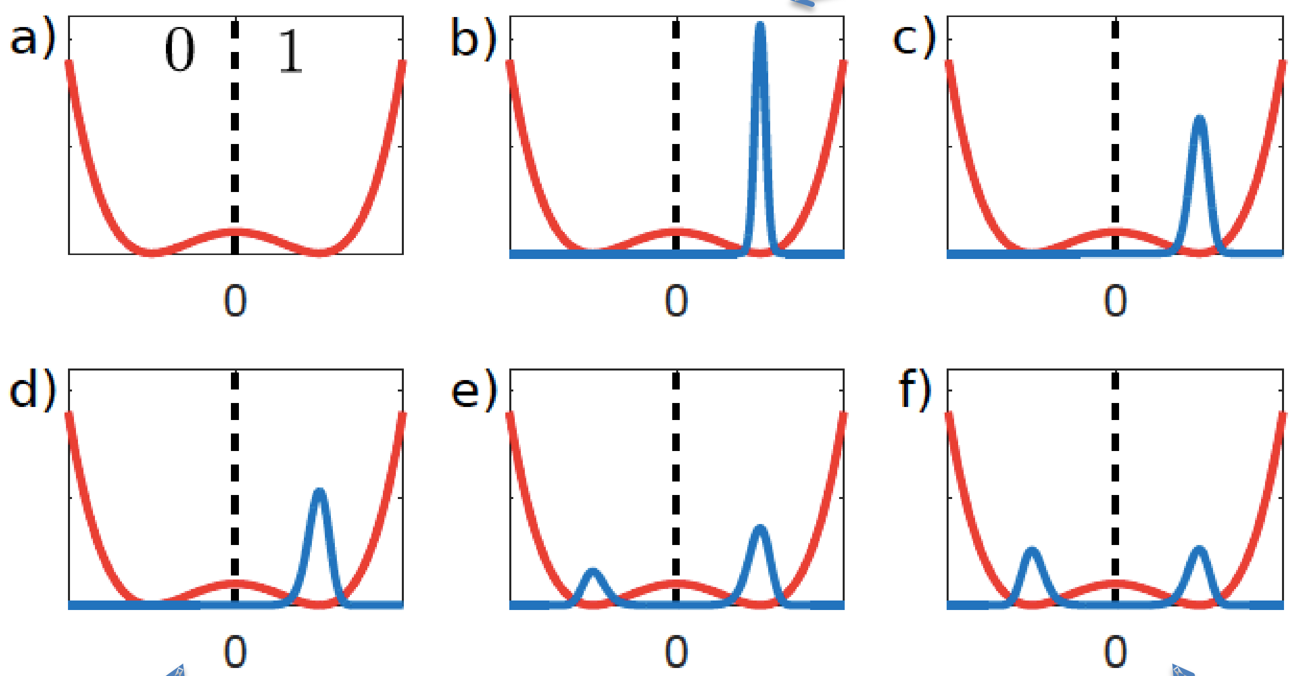


Relaxation within one well
With time scale t_w

Relaxation toward global
Equilibrium with time scale τ_K

1 1-bit memory

Initial probability density p(x,t)



Relaxation within one well
With time scale t_w

Relaxation toward global
Equilibrium with time scale τ_K

Memory lost

2 Compute $\mathbf{t_R}$ for a given set of $\mathbf{P_E}$ and $\mathbf{t}$

In this framework if we indicate with $P_0(t) = \int_{-\infty}^{0} p(x,t)\mathrm{d}x$

the probability to be in the wrong well (bit 0 instead of bit 1), we have:

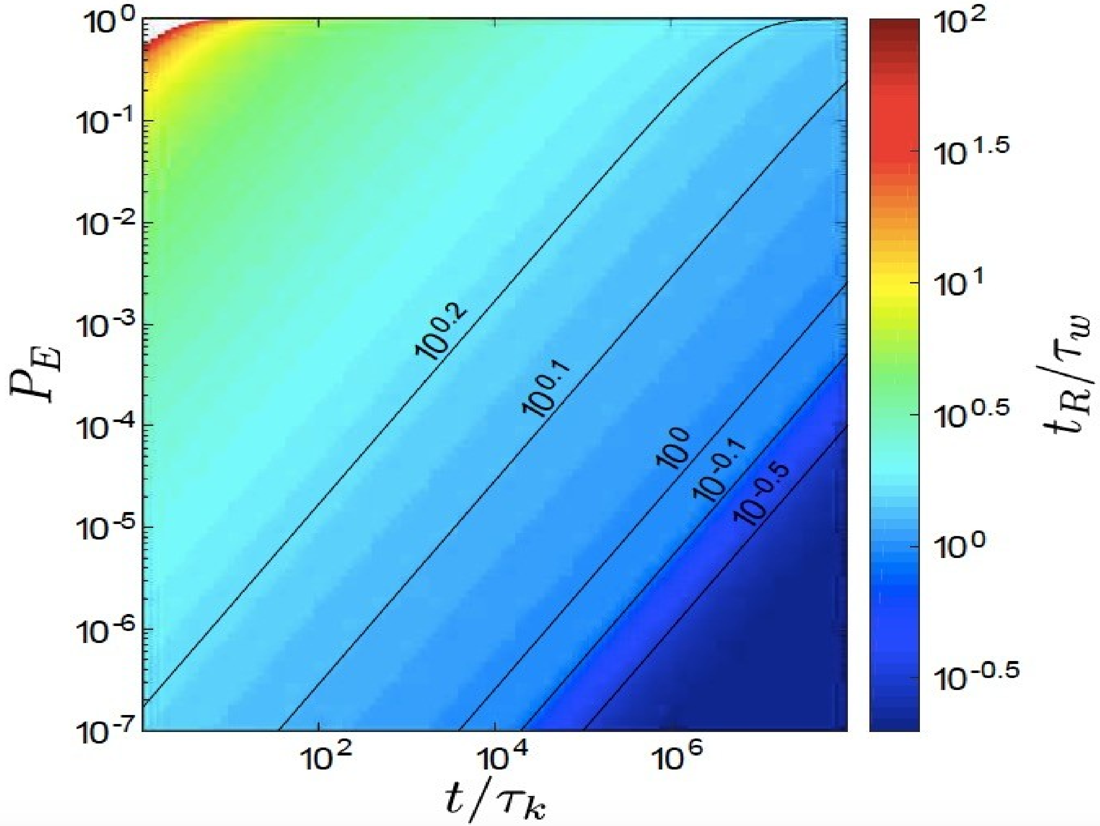
$$P_E = 1 - \left[1 - P_0\left(t_R\right)\right]^{\frac{t}{t_R}}$$

After $N = t/t_R$ refresh cycles

In order to compute this quantity we assume a bistable Duffing potential U(x). The density function p(x, t) is described via the dimensionless Fokker-Plank equation

$$\frac{\partial}{\partial t}\, p(x,t) = \frac{\partial}{\partial x}\left(\frac{\partial U}{\partial x}\, p(x,t)\right) + \frac{k_B T}{\Delta U}\frac{\partial^2}{\partial x^2}\, p(x,t)$$

2 Compute t_R for a given set of P_E and t

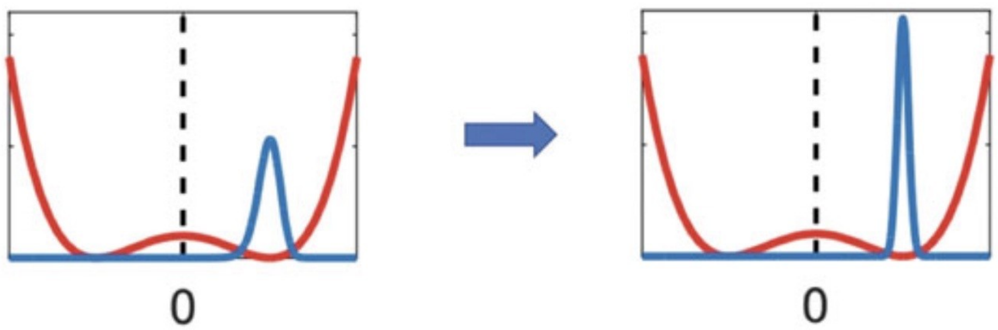


For a given total duration t, the smaller is the acceptable P_E, the more frequent I have to refresh.

3 Determine the minimum energy for keeping the memory

We now consider the energy cost of a single refresh operation.
Based on our model, the refresh operation consists in bringing the p(x,t) back to its initial condition:

$$p(x,t_R) \rightarrow p(x,0)$$



3 Determine the minimum energy for keeping the memory

We assume that the motion inside one well can be approximated by the harmonic oscillator dynamics. This is reasonable while $t_R << \tau_K$.

The resulting probability density function is a sum of two Gaussian peaks centred around the minima of **U**(**x**), each one with the same standard deviation σ

The refresh operation amounts to change $p(x,t_R) \rightarrow p(x,0)$

or $\sigma_f = \sigma(t_R)$ into $\sigma_i = \sigma(0)$

with

$$\sigma(t_R) = \sqrt{\sigma_w^2 + \exp\left(-\frac{t_R}{\tau_w}\right)\left(\sigma_i^2 - \sigma_w^2\right)}$$

3 Determine the minimum energy for keeping the memory

Under this assumption we write its expression for the Duffing potential within the assumed approximation. The minimum energy required is due to the decrease in entropy during the refresh.

MINIMUM ENERGY REQUIRED TO PRESERVE A MEMORY OVER A FIXED TIME WITH A GIVEN ERROR PROBABILITY

$$Q_m = -NT\Delta S = \frac{t}{t_R} k_B T \ln \left(\frac{\sqrt{(\sigma_w^2 + e^{-\frac{t_R}{\tau_w}} (\sigma_i^2 - \sigma_w^2))}}{\sigma_i} \right)$$

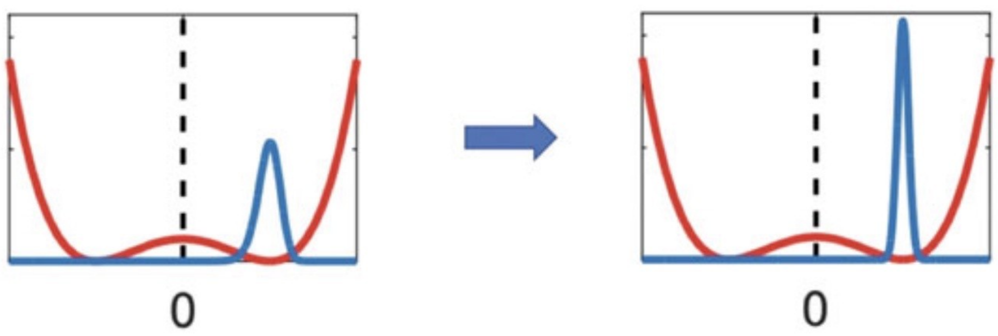
Where: σ_i is the initial probability density width

σ_w is the equilibrium (inside one well) probability density width

3 Determine the minimum energy for keeping the memory

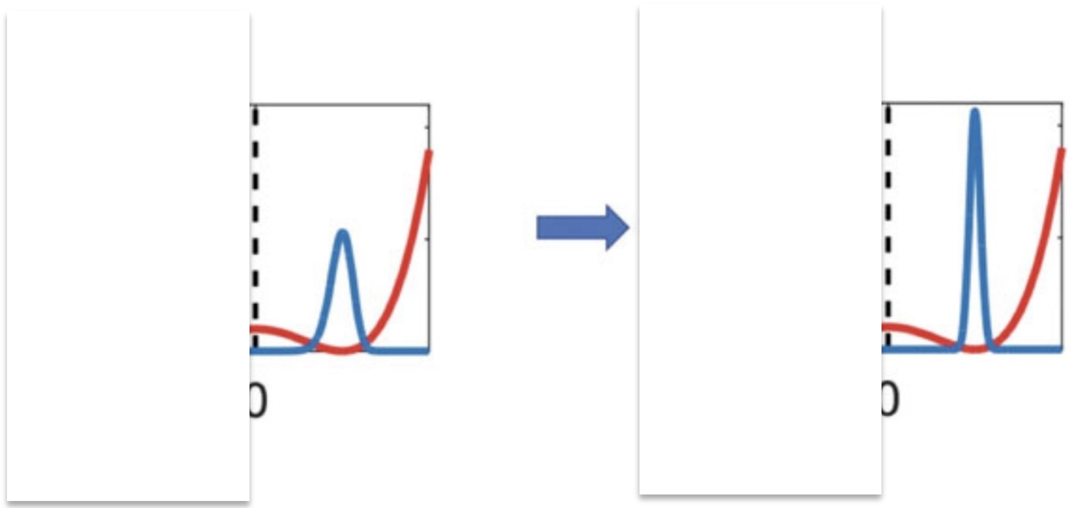
Is this minimum energy physically attainable?

Let's test it with an experiment, where $p(x,t_R) \rightarrow p(x,0)$



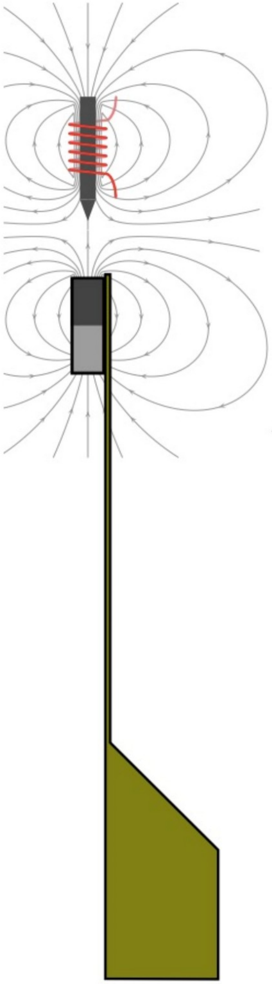
3 Determine the minimum energy for keeping the memory

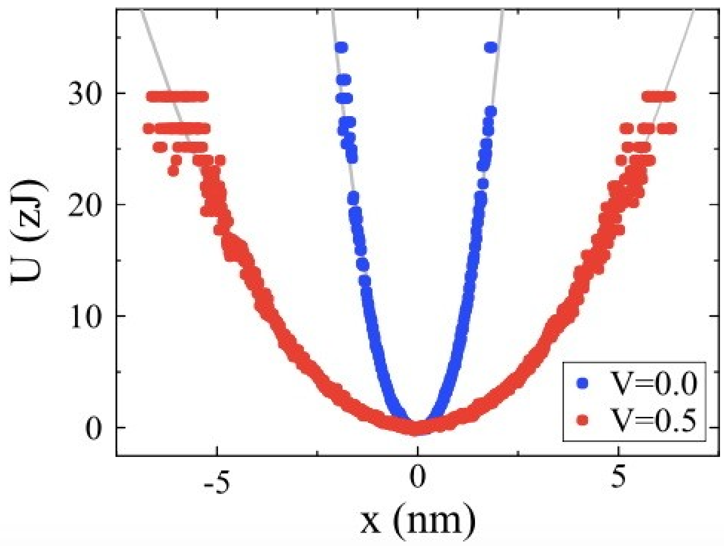
We focus inside a single well and “squeeze” the distribution by modulating the potential

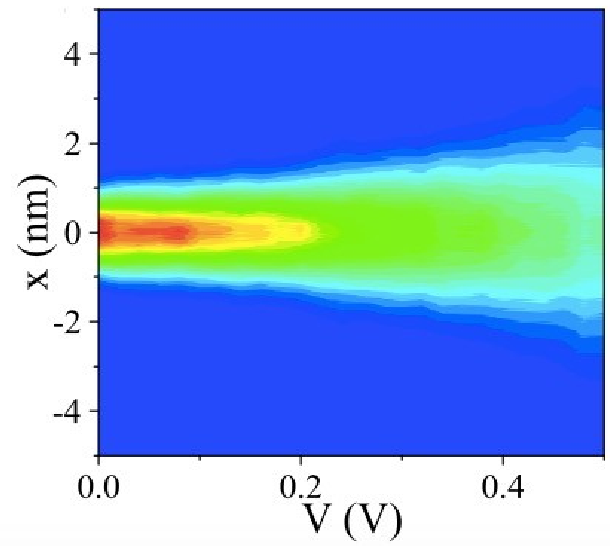


3 Determine the minimum energy for keeping the memory

To"squeeze" the density function inside an harmonic well, we perform an experiment with a micro-mechanical V-shaped cantilever where the relevant observable x is the tip position, by changing the stiffness of the potential.

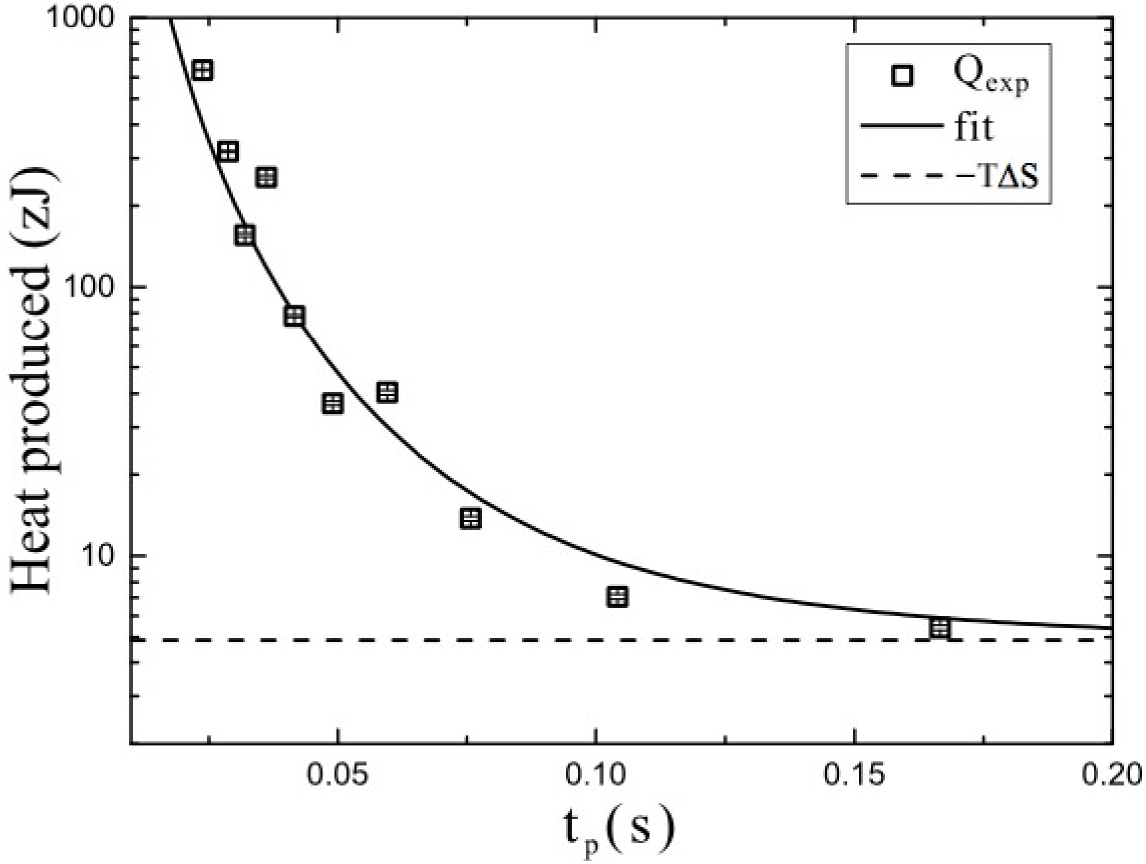






We measure the dissipated energy during the process as a function of the process time

3



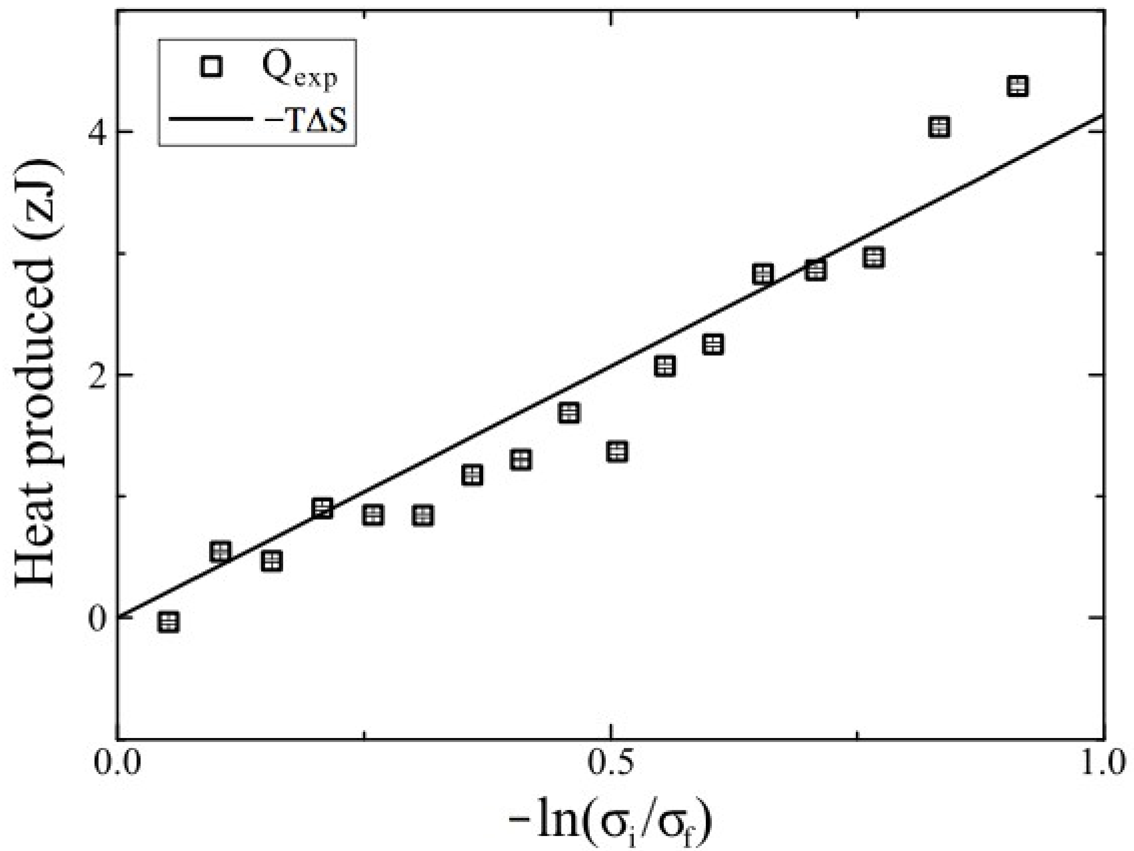
T ΔS

1 zJ = 10^{-21}J

NiPS Laboratory
Noise in Physical Systems

We measure the dissipated energy during the process as a function of the amount of squeezing

3



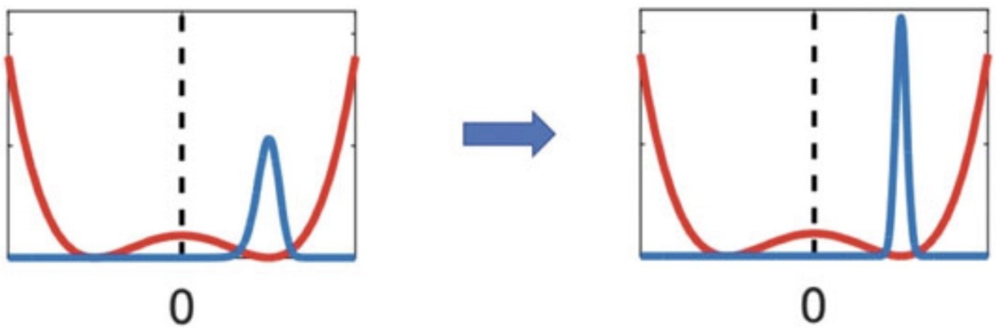
With $\Delta S = k_B \ln\left(\frac{\sigma_i}{\sigma_f}\right)$

3

Going back to the previous question

Is this minimum energy physically attainable?

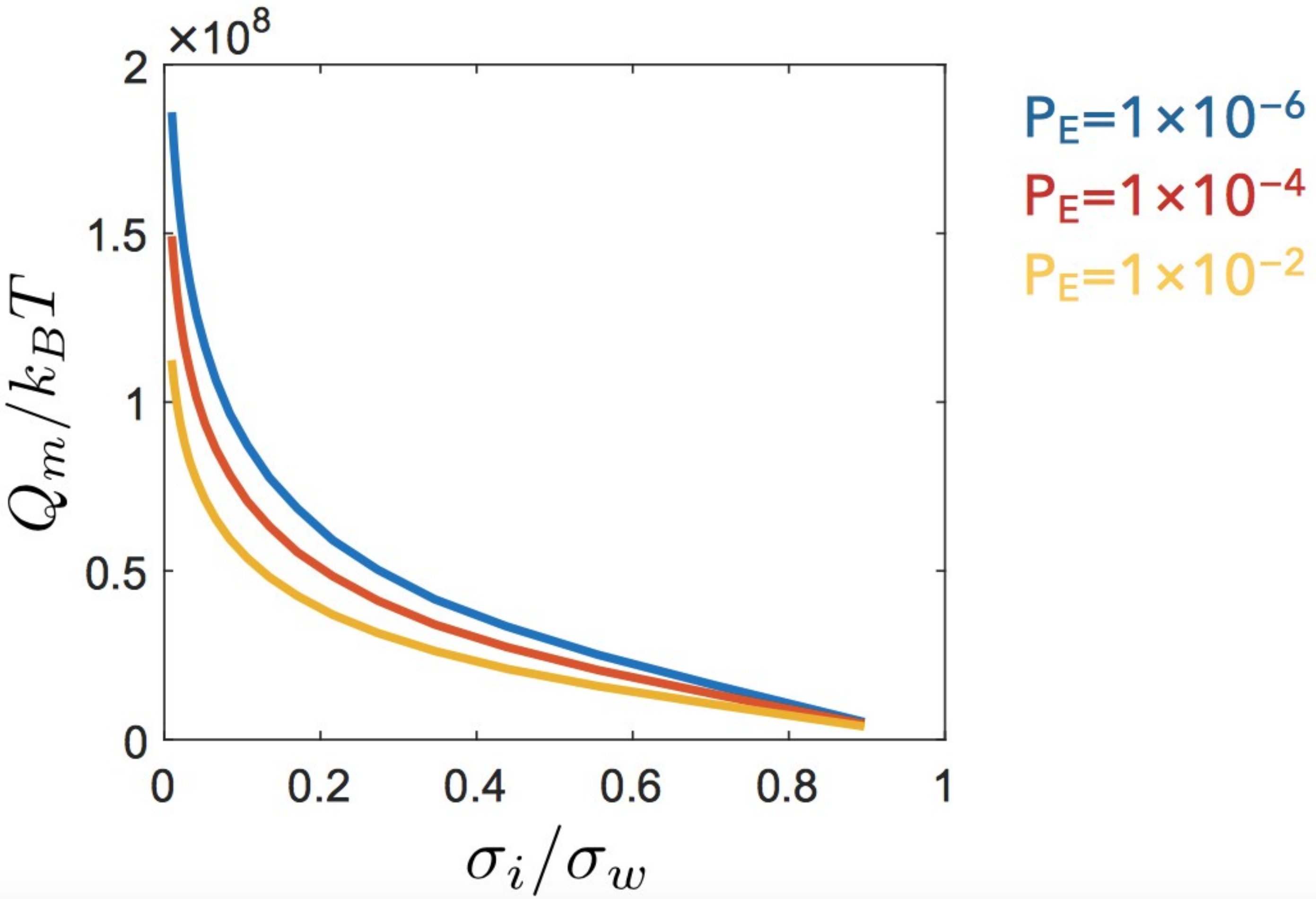
The answer is YES provided the process is performed slow enough



Another question: how small this minimum energy can be made?

3

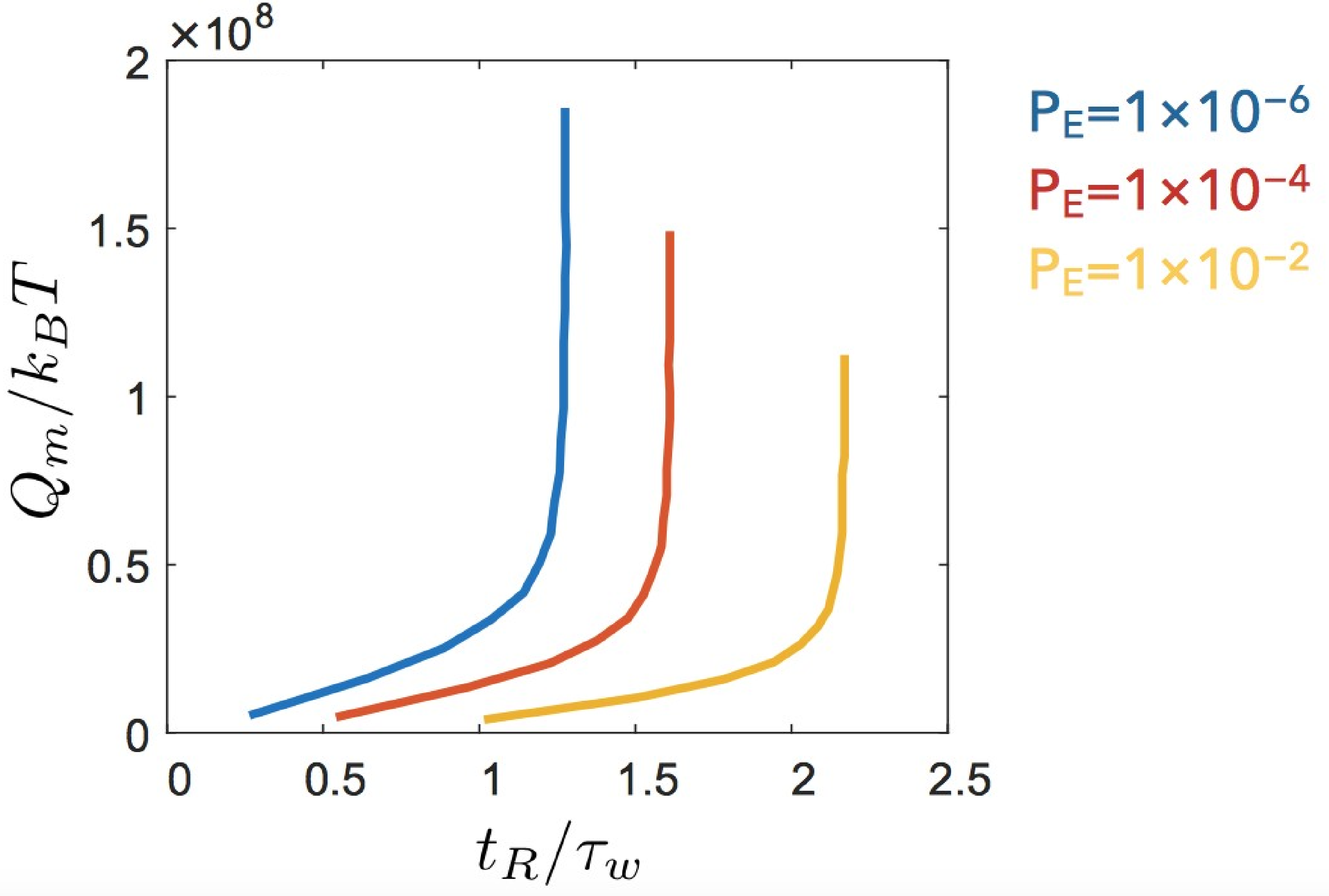
If we keep the system close to equilibrium is better



NiPS Laboratory
Noise in Physical Systems

3

If we refresh very often is better

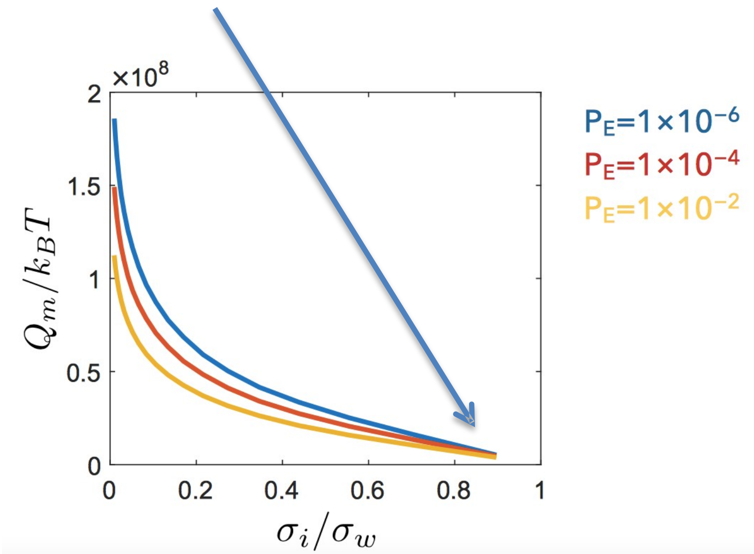


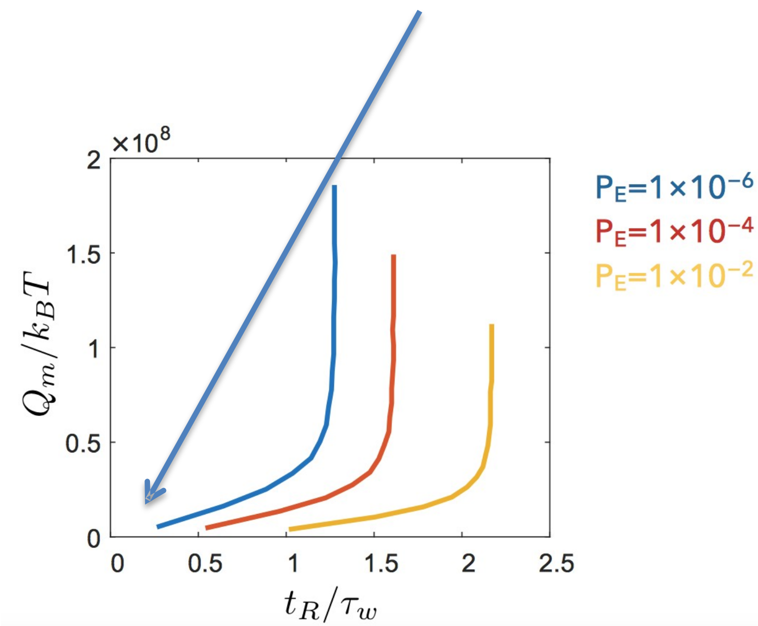
4 Considerations

The good news!

We can preserve a memory for a given time with a given error probability while spending an arbitrarily little amount of energy.

This is accomplished if the refresh procedure is performed arbitrarily often or arbitrarily close to thermal equilibrium.





NiPS Laboratory
Noise in Physical Systems

4 Considerations

The bad news!

If we consider the relation:

$$P_E = 1 - \left[1 - P_0\left(t_R\right)\right]^{\frac{t}{t_R}}$$

we have:

$$t = t_R \ln(1 - P_E)/\ln(1 - P_0)$$

Once we set P_E and select a finite t_R, we can make **t** as large as we want by properly selecting P_0 small enough.

However P_0 cannot be made arbitrarily small without spending a finite amount of energy.
This can be seen in terms of the width σ_0 of the initial distribution.
If we want to make $\sigma_0 = 0$, we need to perform an operation that changes the system entropy form a given σ_i to $\sigma_f = \sigma_0$.
As we have seen, the associated change in entropy is provided by

$$\Delta S = k_B \ln\left(\frac{\sigma_i}{\sigma_f}\right)$$

Thus when $\sigma_f = \sigma_0 \rightarrow 0$ the entropy change tends to diverge and so does the energy required to perform this operation.

4 Considerations

Another way to look at this problem is to consider the **Heisenberg Indetermination principle** that prevents the arbitrary confinement of the probability density, without spending an infinite amount of energy: the uncertainty on the impulse diverges when the uncertainty on the position shrinks.

In the best scenario we have: $$\sigma_x \sigma_p = \frac{\hbar}{2}$$

If the memory setting operation is performed at thermal equilibrium

we have $$\sigma_p = m\sqrt{\langle v^2 \rangle - \langle v \rangle^2} = \sqrt{m k_B T}$$

and thus $$\sigma_x = \frac{\hbar}{2\sqrt{m k_B T}}$$

Thus it exists a σ_{MIN} and it has to be $$\sigma_i \geq \sigma_{iMin} = \frac{\hbar}{2\sqrt{m k_B T}}$$

4 Considerations

Example:

If we assume the distance between the two wells $\mathbf{x}_m$ = 1 nm and a refresh period $\mathbf{t}_R$ = 6.6 10^{-3} s, we have that the minimum σ_i = 9. 6 10^{-20} m.

For $\mathbf{P}_E$ = 1 10^{-6} then the maximum value for **t** is approximately 2 years.
For $\mathbf{P}_E$ = 1 10^{-4} then the maximum time **t** is approximately 200 years.

4 Considerations

The existence of a σ_{MIN} implies that the probability of error P_0 cannot be arbitrarily small

and, thus $t = t_R \ln(1-P_E)/\ln(1-P_0)$ cannot be arbitrarily large

For any P_E we select we have an associate maximum for the memory duration **t**

The good news: You can keep your memory by spending 0 energy

The bad news: A memory cannot last forever

Conclusion about the energy of memory preserving

Take home message

You can preserve your memory only for a limited amount of time.

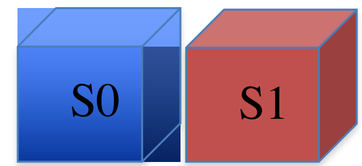
Within this limit, if you do things carefully enough, you do not need to spend any energy.

The cost of remembering one bit of information
Davide Chiuchiù, Miquel López-Suárez, Igor Neri, Maria Cristina Diamantini, Luca Gammaitoni.
Physical Review A 97 (5), 052108, 2018

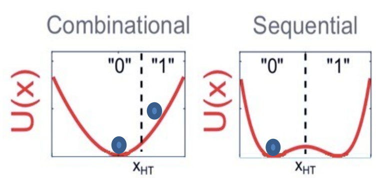
Summary

Computing is messing up with quantities

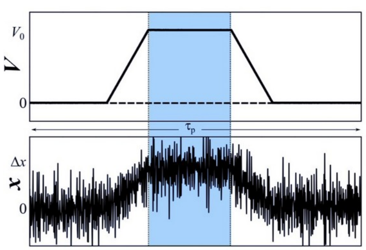
It is better done with the help of machines



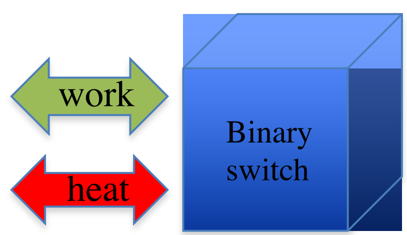
Simplest computing machine



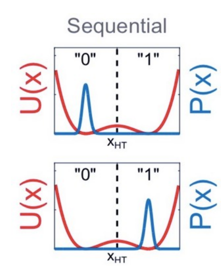
They follow the laws of physics



The presence of fluctuations require small scale thermodynamics



So that we have to take into account Heat and Entropy



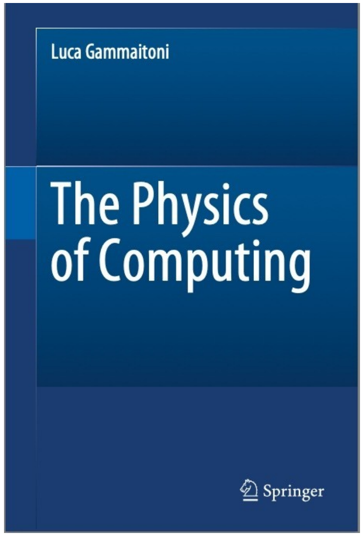
In the framework of non equilibrium statistical mechanics

Summary

We have addressed the following questions, associated to fundamental limits is the functioning of computing devices (logic gates and memories):

- What is the minimum energy we have to spend if we want to produce a switch event ? **ZERO energy**
- Does this energy depends on the technology of the switch ? **NO it is a fundamental limit**
- Does this energy depends on the instruction that we give to the switch ? **NO provided the switch is adiabatic**
- How much energy do we need to spend if we want to memorize data ? **A minimum of K_B T log 2**
- How much energy do we need to spend if we want to keep the memorized data ? **ZERO energy but it will not last forever**

To know more



L. Gammaitoni, The Physics of computing, Springer, 2021