Thermodynamics of computation: a summary of very classic and fundamental issues.

In this lecture we go through classic material on thermodynamics of computation. Thermodynamics was born in the 19th century through questions intimately related to thermal engines transforming heat into mechanical work and vice-versa. A central concept which emerged is "entropy" which now pervades many fields beyond classical thermodynamics. (See for example en.wikipedia.org/wiki/Entropy.)

Thermodynamics of computation similarly concerns "modern machines" which are "computing devices" & "sensing devices". The basic questions are again: what are the relations between heat, work, and the task of information processing. These questions are relevant both for classical information processing and quantum information processing. There is already a long history of investigations in this realm.

Today we touch upon very classic material:

1) Maxwell's demon
2) Szilard's engine
3) Landauer's principle
4) Bennett's view on Maxwell demon & Szilard's engine.
   Modern solution of paradoxes through Landauer's principle

5) Irreversible and reversible computing.

6) A deeper dive into Landauer's principles:

   Study of two models in the litterature.

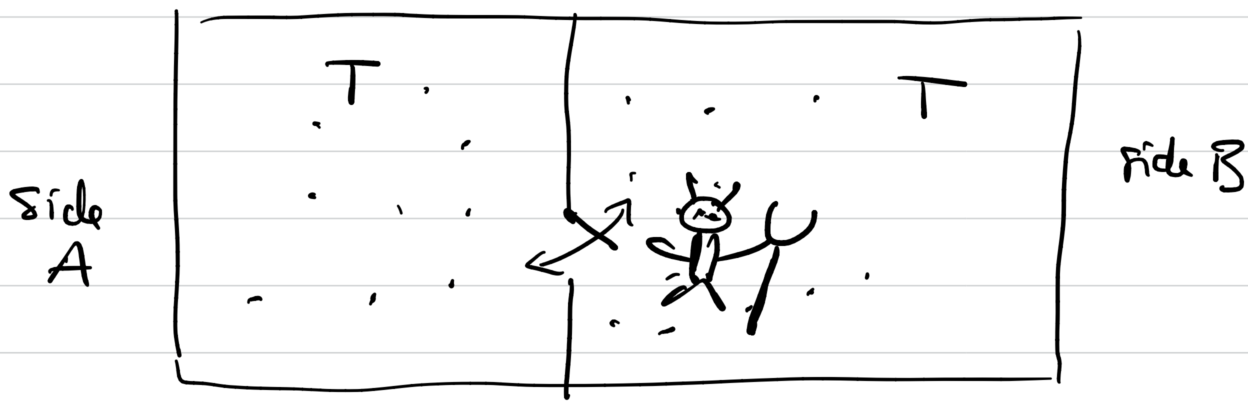
1) Maxwell demon.

In 1867 Maxwell proposed the following thought experiment in a letter to lord Kelvin (it is Kelvin who coined the name Maxwell demon).

Consider a gas at thermal equilibrium at temperature T in a container separated by a barrier. A demon with supernatural powers controls a small door between the two chambers and can decide to let pass or not molecules from one side to the other. More specifically the velocity distribution of molecules is

$$\rho(\vec{v})\, d^3\vec{v} = \left(2\pi \frac{k_B T}{m}\right)^{-3/2} \exp\left(-\frac{1}{2}\frac{m|\vec{v}|^2}{k_B T}\right).$$

The average velocity vanishes $\langle \vec{v} \rangle = 0$ and the mean kinetic energy $\langle \frac{1}{2} m \vec{v}^{\,2} \rangle = \frac{3}{2} k_B T$. Some molecules are therefore slow and some are fast.



The demon acts as follows:

- Lets through molecules from B → A if they have high velocity
- Lets through molecules from A → B if they have low velocity
- Blocks molecules from passing B → A if they have low speed
- Blocks molecule from passing A → B if they have high speed.

Eventually the gas ends up in a situation with high speed molecules on the left and low speed molecules on the right. Correspondingly the temperatures of the A and B side satisfy $T_A > T_B$

since $\frac{3}{2} k_B T_A = \langle \frac{1}{2} m \vec{v}_A^{\,2} \rangle \geq \langle \frac{1}{2} m \vec{v}_B^{\,2} \rangle = \frac{3}{2} k_B T_B$.

$T_A > T_B$

The demon appears to violate the second law of thermodynamics (*). Indeed the initial situation is a higher entropy state than the final situation. Moreover it would be possible to use the gas in the final state to extract mechanical work from the two reservoirs at different temperatures. However this paradoxical because overall this means that we extracted work from the single reservoir at equilibrium temperature T.

(*) We assume that the movements of the trap door are so small that the work expenditure of the demon are negligible

It is also easy to think of variations of this thought experiment where the demon would set a heat flow from a cold reservoir to a hotter one; which again violates the second law of thermodynamics if this demon does not expend work to open and close the little trapdoor.

2) Szilard's engine.

In 1929 Szilard proposed a reformulation of Maxwell's paradox and made a detailed analysis in order to 'solve' it. In a nutshell Szilard argued that the demon needs to <u>measure</u> the speed of the molecules incoming near the trapdoor, and that this measurement process will dissipate heat and expend work. Followers (Gabor, Brillouin, Feynman, ...) also explained the paradox in similar terms.

As we will see later Bennett later criticized this explanation as being really fundamental and proposed another resolution based on Landauer's principle (see later).

For the moment lets summarize here the essentials of Szilard's analysis.

Szilard engine is a cyclic process as follows.

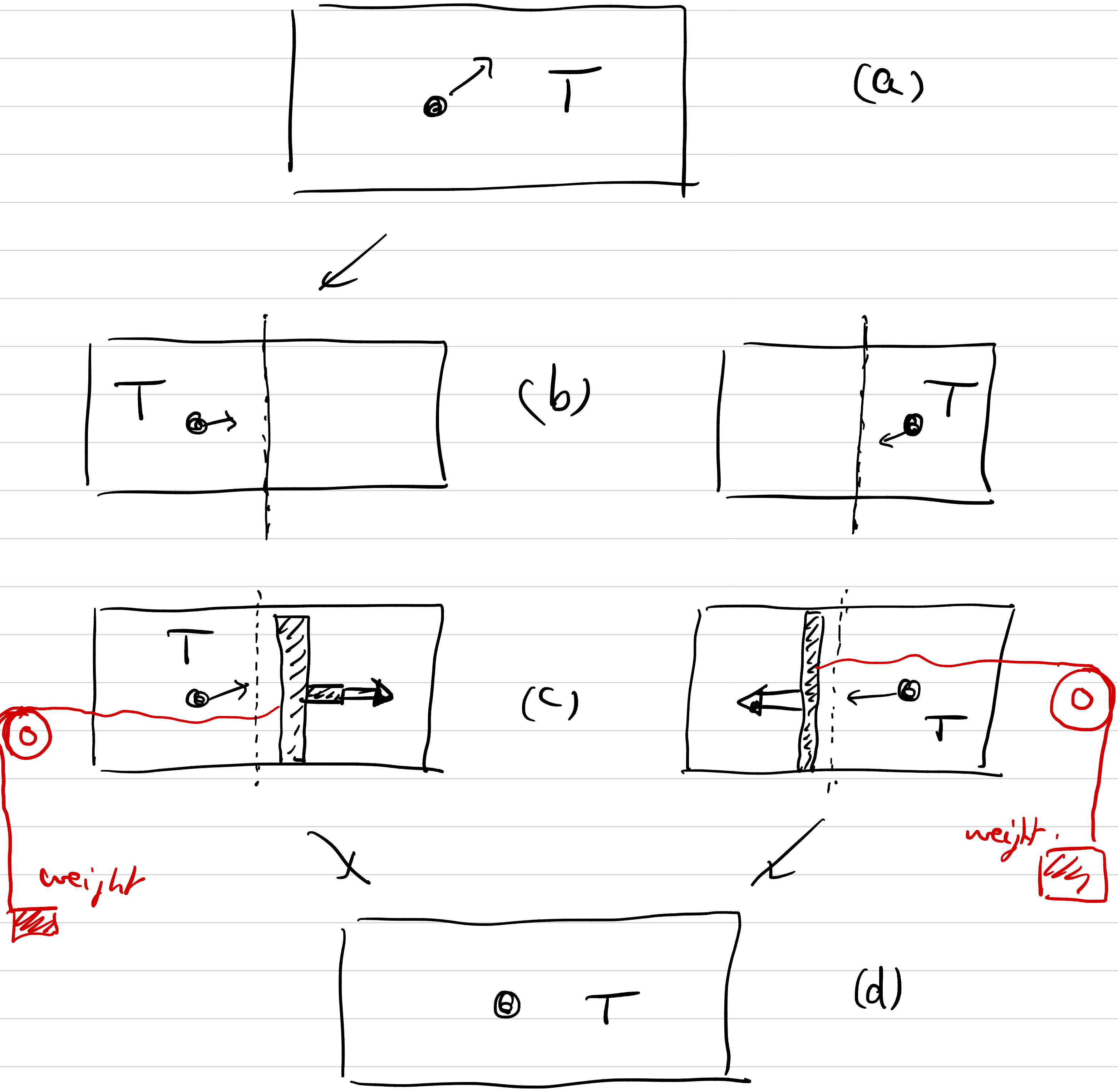
a) A (single) molecule is in a box at thermal equilibrium (temperature T) with its environment (outside the box).

b) We insert a barrier in the middle of the box.

c) We measure location (left/right) of the molecule with some measurement device. If molecule is on the left we insert a piston on the right, remove the barrier, and slowly let the piston move to the right in an isothermal way (temperature T remains unchanged).

If molecule is on the right, we insert a piston on the left, remove the barrier, and slowly move the piston to the left in an isothermal way.

d) At the end of this cycle the molecule again occupies the whole box. We can repeat this cycle endlessly.



In step (c) the gas exerts mechanical work on the environment (as exemplified by a weight that could be lifted). This is paradoxical as it is a cyclic process that can be repeated again and again to extract work (from box to environement) although the system is at thermal equilibrium at one single temperature T. We have here a perpetuum mobile!

Before discussing Szilard's proposal to resolve the paradox, let us compute the amount of work extracted.

Since the process is isothermal we have that the pressure exerted on the piston satisfies $pV = N k_B T$ with here $N = 1$. So

$$p = \frac{k_B T}{V}.$$

Work for displacement dx is $\underbrace{\text{Force}}\cdot dx =$

$$p S dx = p dV$$

(with S = transverse surface of piston).

Thus along one stroke of the cyclic engine:

$$\Delta W = - \int_{V/2}^{V} \frac{k_B T}{(\mathcal{S}\ell)}\, d(\mathcal{S}\ell)$$

$$= - k_B T \left(\ln V - \ln \frac{V}{2} \right)$$

$$= - k_B T \ln \left(\frac{V}{V/2} \right) = - k_B T \ln 2.$$

Here the minus sign means that work is a flux of energy that goes from box to environment (through the piston). So it is "given to".

At the same time we can see that heat must flow from the environment to the box. Indeed from the first law of thermodynamics (energy conservation)

$$\Delta U = \Delta Q + \Delta W$$

where ΔQ = heat flow in box from env
ΔW = work flow from box to env

$\Delta U = 0$ because the process is cyclic and the total internal energy in box is conserved.

$$\Rightarrow \quad \Delta Q = -\Delta W = +\, k_B T \ln 2 .$$

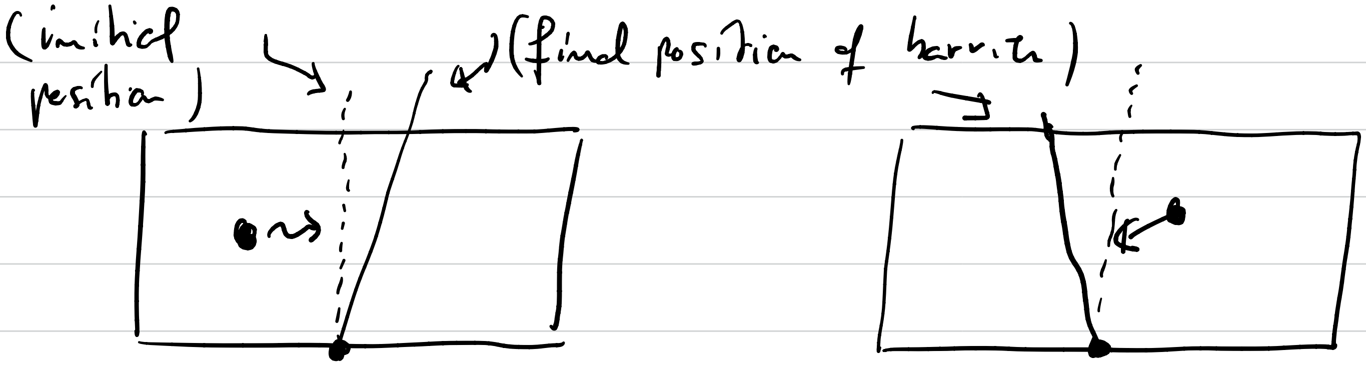
In fact it is this heat flow that maintains the temperature constant as the box expands in the isothermal process.

What is Szilard's solution of the paradox?

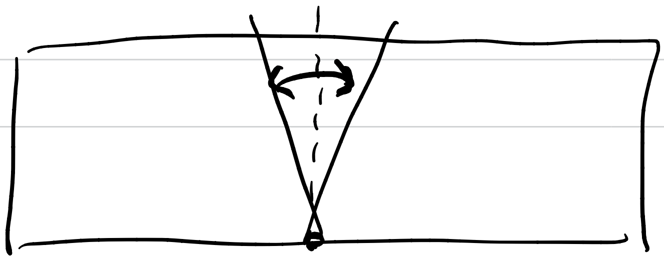
We have only taken into account step (c) so far in the analysis. In fact the measurement process in step (b) should dissipate heat and this will impede eventually the extraction of work as we repeat the cycle. Here we limit ourselves to an intuitive explanation.

In order to get a define value of the location measurement (Right / Left) or (1 / 0) we must have some sort of friction involved.

Imagine a barrier that can turn around some axis and records the Right / Left location of the molecule as an "angle variable":



The friction enables the barrier to get 'stuck' in its final position, thus indicating the side of the box where the molecule resides. However eventually this friction will generate heat and the barrier rotations will thermalize with the reservoir producing



random oscillation and random angles. Hence we will make eventually mistakes on the location of the molecule & on the side in which the piston should be inserted. This will eventually on the long run lead to restitute work from the environment to the piston or box (the weight goes down instead of up). On average the work extracted (and height of weight) stays zero.

A very detailed discussion can be found in Feynman's lectures (vol I) on his famous rachet & pawl version of Szilard's engine.

<u>Critisism of this explanation</u>: in fact as discussed by Bennett one can devise measurement devices that do not involve friction. See Scientific American paper by Bennett for example. More on this later.

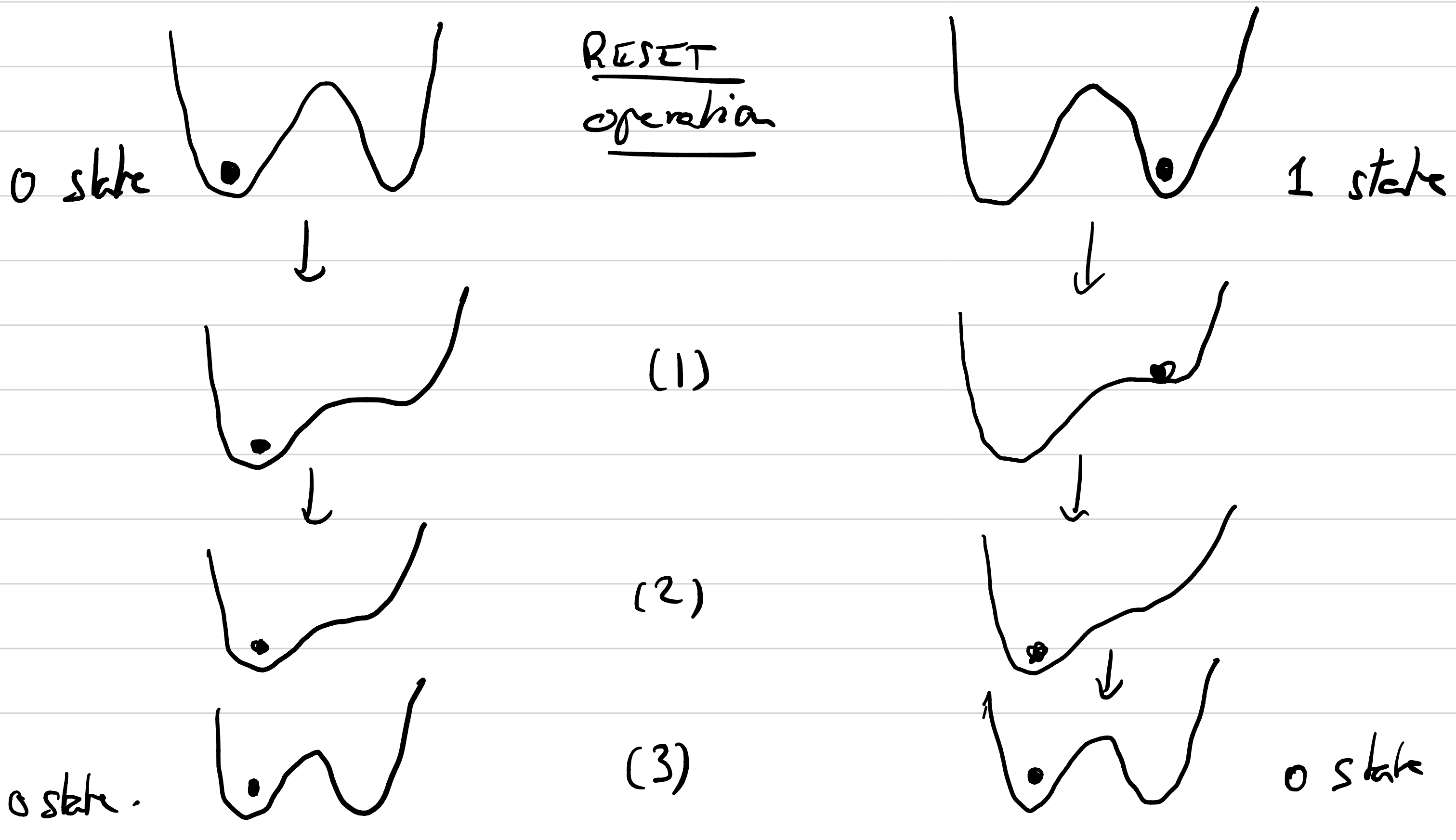
3) Landauer's principle. (1961)

In a quite independent development Landauer (then at IBM research) reflected on the heat dissipated by computing devices. This in turn had a bearing on the modern resolution of the above paradoxes connected with Maxwell's demon and Szilard engine.

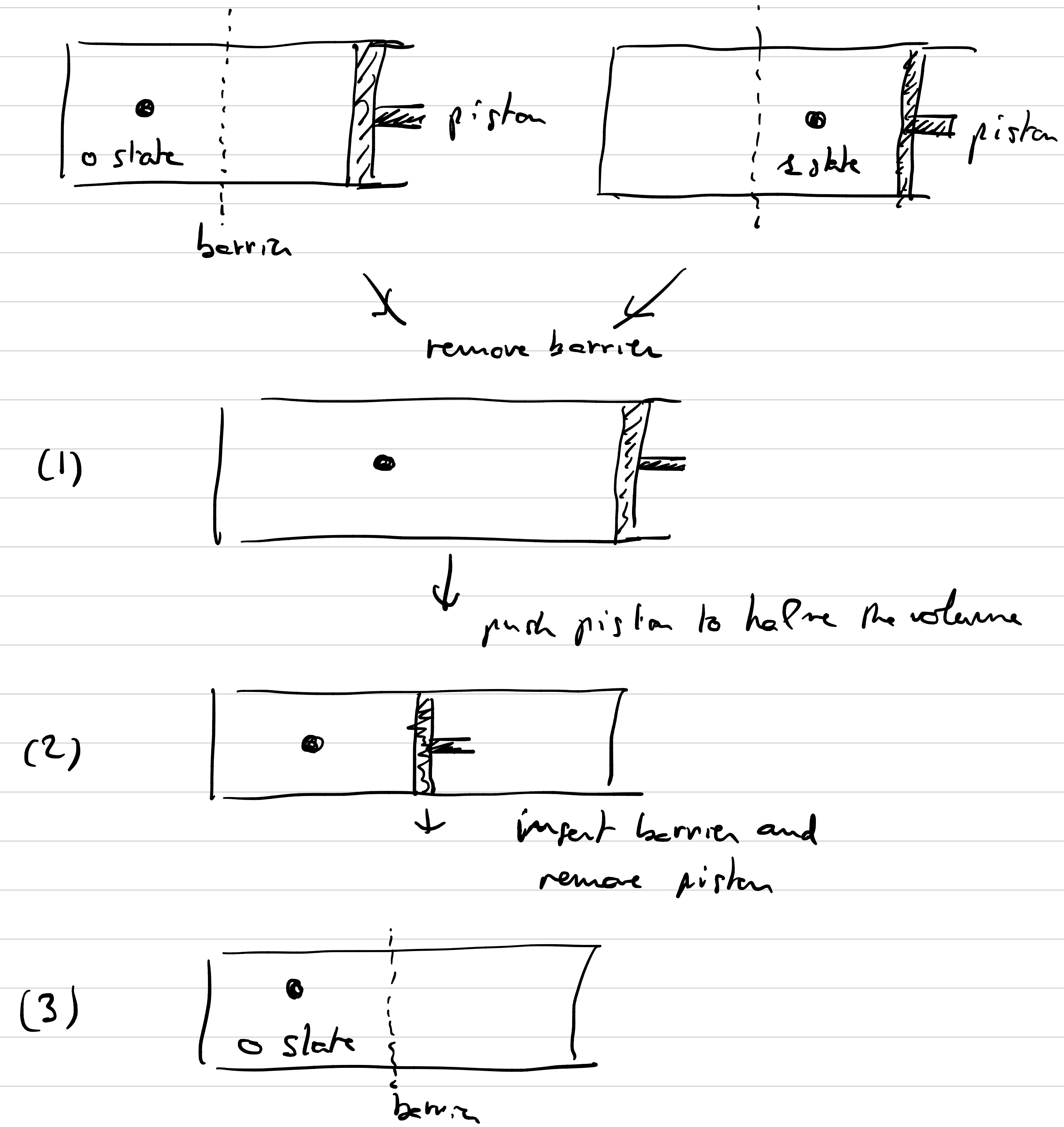
Consider the most elementary information processing task: erasure of 1 bit of information. How to formalise this?

According to Landauer the erasure of 1 bit in a memory amounts to a RESET OPERATION that resets the memory to the ZERO STATE 0 whatever was the initial state (assume the initial state was 0 or 1). One may think of resulting magnetization domains in disks etc... Here is an illustrations used by Landauer:

The memory is a double well potential and the RESET (erasure) operation amounts to apply a time dependence to this double well. The particle goes to the zero state where it stays so due to friction to avoid oscillations.



Similarly one may consider a box model similar to Szilard's engine – but different – here it is a memory device:



All this is just analogous to the double well.

It should be clear from the box model that the piston exerts a work on the 'box' or 'molecule' $\Delta W = + k_B T \ln 2$.
This can be proven just as in the analysis of the Szilard engine (the compression is isothermal) and here the plus sign is because the flux of work/energy goes from environement into the box.
As a consequence by energy conservation a heat flux goes from the box into the environmement and an amount of heat $\Delta Q = - k_B T \ln 2$ is released into the environement.

Perhaps more generally we can also use the following information theoretical argument. The entropy of the initial state of the memory is (with $p_{L/R}$ = prob to find mol in L/R)

$$-p_L \ln p_L - p_R \ln p_R = -\frac{1}{2}\ln\frac{1}{2} - \frac{1}{2}\ln\frac{1}{2}$$

$$= -\ln\frac{1}{2} = +\ln 2.$$

After the reset operation we have $p_L = 1$, $p_R = 0$

thus the entropy is 0.

$$(\Delta S)_{Box} = S_{final} - S_{initial} = 0 - \ln 2 = -\ln 2$$

Using $\quad (\Delta S)_{Box} = \dfrac{(\Delta Q)_{Box}}{T}$ we find

$(\Delta Q)_{Box} = -k_B T \ln 2$ which flows in the box.

$(\Delta Q)_{Env} = +k_B T \ln 2$ flows in the environment.

We arrive at Landauer's principle:

> "The erasure of 1 bit of information necessarily is accompagnied by at least $k_B T \ln 2$ of heat released in the environement".

There have been many debates and critisisms about the general validity of this claim. Various studies have analyzed models to prove the claim. Of course these proofs are always valid within the models. We give two interesting references that 'prove' the claim for stochastic, deterministic, and quantum dynamical models:

* K. Shizume Phy Rev E 52, 3495 (1995)
* B. Piechocinska Phys Rev A 61, 062314 (2000).

4) Modern resolution of Maxwell's demon and Szilard's engine paradoxes

The modern solution of the paradoxes came quite late in the 20th century (around 1980's). The solution offered by Szilard and followers can be critised because it is in principle possible to perform measurement in a dissipation less way (i.e without 'friction'). In fact Bennett and other have provided examples of mechanical device that do so (see for example Scientific American review of Bennett 1987). A conceptual/formal idea is as follows. Let $x \in \{0, 1\}$ a degree of freedom to be measured (e.g L/R location of molecule). The measurement means that we record the value of this degree of freedom in a variable y. Mathematically this is a CNOT operation:

$$\mathrm{CNOT}(x, y) = (x, y \oplus x)$$

with y set to a predefined state (say $y = 0$). This operation is logically invertible. There is no loss of information and therefore it should be possible to implement it with a device that operates in a thermodynamically reversible way without entropy loss or heat dissipation.

But then how do we resolve the paradoxes of Maxwell's demon and Szilard's engine?

According to Bennett and Landauer in order to repeat the cycle of the engine the value $y \oplus x$ recorded in the "demon's memory" has to be erased. This should be clear if we assume the memory is finite. (If it is infinite this poses other conceptual problems as the process is then not cyclic and never ending; we do not discuss this issue).

The erasure of the demon's memory will, according to Landauer's principle, dissipation an amount of heat ($k_B T \ln 2$) from the memory to the environement. At the same time this requires expenditure of work $\Delta W = k_B T \ln 2$ to be done on the memory.

5) Irreversible versus reversible computation.

As we all know our computing devices dissipate heat. Landauer's principle discussed before can be extended to logical operations that compress the "phase space"; or in other words are not "invertible". These satisfy $\Delta S = S_{final} - S_{initial} < 0$ and we expect a minimal amount of heat $\Delta Q = k_B T \Delta S$ will be dissipated.

It is accepted today that Landauer's principle sets a limit on the minimum possible amount of heat dissipated by devices that process information in logically irreversible ways. (The map output $= \phi$ (input) is not invertible).

For example AND, OR, NAND, NOR are all irreversible.

$$AND(x,y) = x \wedge y = xy$$

$$OR(x,y) = x \vee y = 1 \oplus \bar{x}\bar{y}$$

$RESET(x) = 0$ is also irreversible.

Assuming that the erasure or RESET operation can be used to set the order of magnitude of the minimum heat dissipated in a computation we find $k_B T \ln 2 = 2{,}9 \times 10^{-21}$ Joules

$$= 0{,}018 \text{ eV}$$

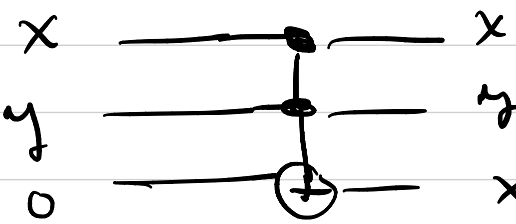
per "operation" at room temperature. Todays computers dissipate much larger amounts of heat of course. Koomey's law states that every 1,57 years (more recent estimate give 2,6 years) the energy consumption of our devices is divided by 2. (This is an analog of the famous Moore law which applies to physical size instead of energy consumption).

According to Koomey's law we would be reaching Landauer's limit in approx 2080 or so.

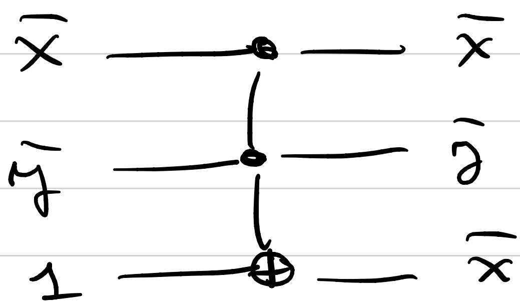
Can one do better and compute without heat dissipation?

According to Bennett (and Toffoli, Fredkin, and others) this is possible by using logically reversible gates. Logically reversible gates ought to be implementable by devices that do not, in principle, dissipate heat since phase space is not compressed. This is briefly discussed in the notes of CS-308 and we will not go in great detail here.

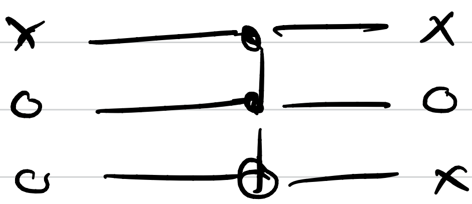
The main idea is to replace the AND, OR, COPY gates by Toffoli gates as follows;



$xy = x \wedge y$
$= \text{AND}(x, y)$



$\bar{x}\,\bar{y} \oplus 1 = \text{NOT}(\bar{x} \wedge \bar{y})$
$= x \vee y = \text{OR}(x, y)$



$x = \text{COPY}(x).$

By a theorem of Emil Post (beginning of 20th century) any binary input - binary output function can be computed from an acyclic circuit composed of (AND, OR, COPY) (or NAND, COPY). Thus using Toffoli gates and extra bits the circuit can be made logically reversible.

There is no general physical principle known that requires heat dissipation in a logically reversible computation.

More on this can be found in Bennett review paper (see moodle page).

6) A deeper dive into Landauer's principle

We propose two papers from the litterature that model in quite detail the steps of Landauer's analysis.

Perhaps the most faithful to Landauer's original arguments is the one by K. Shizume Phys Rev E 52, 3495 (1995). This is based on the analysis of a stochastic process involving friction to perform the RESET operation properly. This study concerns a purely classical bit.

In B. Piechocinska Phys Rev A 61 062314 (2000) a more general argument is given which applies to classical as well as quantum models. The heat bath however is modelled in a purely Hamiltonian way which makes the argument not entirely complete.

Here we present an analysis of Landauer's principle, following B. Piechocinska for a toy quantum bit. We model the RESET operation (erasure) for one quantum bit.

A quantum toy model illustrating Landauer's principle.

The qubit memory is modeled as a two level system with degenerate energy levels

$$H = \begin{pmatrix} \varepsilon_0 & 0 \\ 0 & \varepsilon_0 \end{pmatrix}$$

corresponding to states $|i\rangle \in \{|0\rangle, |1\rangle\}$.

The initial state of this memory is

$$\rho_{init} = p_0^{init} |0\rangle\langle 0| + p_1^{init} |1\rangle\langle 1|$$

The entropy is maximal if $p_0^{init} = p_1^{init} = \frac{1}{2}$.

The RESET operation of the memory should bring the state to

$$\rho_{final} = |0\rangle\langle 0| .$$

The qubit is coupled to a heat bath (e.g. a photon gas) here simply modeled by a thermal density matrix

$$\rho_{Res} = \sum_m \frac{e^{-\beta E_m}}{Z} |m\rangle\langle m|$$

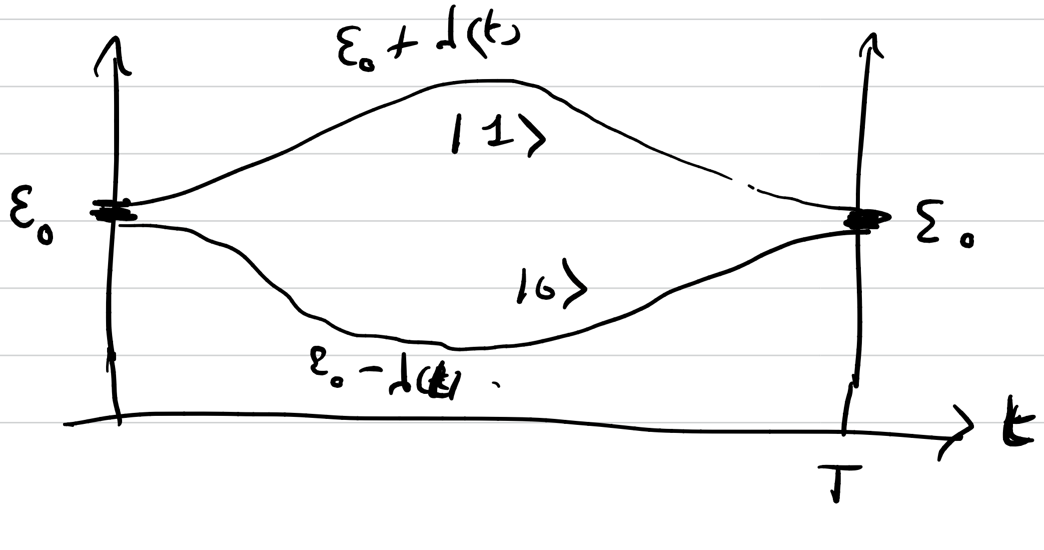
$$Z = \sum_m e^{-\beta E_m} \quad , \quad \beta = (k_B T)^{-1} .$$

The RESET operation proceeds as follows.

We apply a time dependent field to the qubit

$$H(t) = \begin{pmatrix} \varepsilon_0 & 0 \\ 0 & \varepsilon_0 \end{pmatrix} + \lambda(t) \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}_{\sigma_z}$$

which splits the spectrum as follows:



with $d(0) = d(T) = 0$.

- If the memory start in state $|1\rangle$ after some time it will release a photon and end in state $|0\rangle$. This is due to a coupling between the qubit and reservoir.
- If the memory starts in state $|0\rangle$ we suppose that it stays in state $|0\rangle$. This requires that for most of the time $d(t) > \max_m E_m$ so the qubit cannot absorb a photon.
- Finally we suppose that the reservoir (heat bath) is subject to some decoherence so that after the photon has been emitted its density matrix stays diagonal in the energy basis $|m\rangle$.

Thus if the initial state of the reservoir was $|m\rangle$ and the final state $|m'\rangle$, the energy flux or heat dissipated by the memory into the reservoir is

$$\Delta Q = E_m - E_n \quad .$$

We stress that these three points are quite heuristic and a more deep treatment would be needed (used some open quantum system formalism such as Lindblad dynamics).

Consider now a transition between initial and final state of (M = memory ; R = reservoir)

$$|i\rangle_M \otimes |n\rangle_R \longrightarrow |f\rangle_M \otimes |m\rangle_R$$

Set

$$\Gamma_{in;fm} = \ln p_i^{init} - \ln p_f^{final} + \beta(E_m - E_n)$$

For the moment $p_i^{init} = (p_0^{init}, p_1^{init})$ and $p_f^{final} = (p_0^{final}, p_1^{final})$ are general prob distr.

Latter we specify them for the RESET operation to $\rho_0^{init} = \rho_1^{init} = \frac{1}{2}$ & $\rho_0^{final} = 1$, $\rho_1^{final} = 0$

In $\Gamma_{i_m; f_m}$ the indices are random variables with joint probability distribution

$$|\langle f, m \,|\, \underbrace{U_T}_{\text{'unitary evolution'}} \,|\, i, m\rangle|^2 \, p_i^{init} \cdot \frac{e^{-\beta E_m}}{Z}.$$

Therefore

$$\mathbb{E}\left(e^{-\Gamma_{i_m f_m}}\right) = \langle e^{-\Gamma_{i_m f_m}} \rangle$$

$$= \sum_{i_m f_m} e^{-\Gamma_{i_m f_m}} |\langle f m | U_T | i m\rangle|^2 \rho_i^{init} \frac{e^{-\beta E_m}}{Z}$$

$$= \sum_{i_m f_m} \frac{\rho_f^{final}}{\cancel{p_i^{init}}} e^{-\beta(E_m - \cancel{E_m})} |\langle f m | U_T | i m\rangle|^2 \cancel{\rho_i^{init}} \frac{\cancel{e^{-\beta E_m}}}{Z}$$

$$= \sum_{i_m f_m} \rho_f^{\text{final}} \frac{e^{-\beta \epsilon_m}}{Z} |\langle f_m | U_T | i_m \rangle|^2$$

$$= \sum_{f_m} \rho_f^{\text{final}} \frac{e^{-\beta \epsilon_m}}{Z} \underbrace{\sum_{i_m} |\langle f_m | U_T | i_m \rangle|^2}_{1}$$

by unitarity of U_T, the lines & columns are normalized to 1.

$$= \underbrace{\sum_f \rho_f^{\text{final}}}_{1} \cdot \underbrace{\sum_m \frac{e^{-\beta E_m}}{Z}}_{1} = 1.$$

$$\Rightarrow \boxed{\langle e^{-\Gamma_{i_m f_m}} \rangle = 1.}$$

Landauer's principle will follow from this formula as we now show.

Using convexity of the exponential

$$\langle e^{-x} \rangle \geq e^{-\langle x \rangle}$$

Jensen inequality.

we find:

$$1 = \langle e^{-\Gamma} \rangle \geq e^{-\langle \Gamma \rangle}$$

$$\Rightarrow \quad \boxed{\langle \Gamma \rangle \geq 0.}$$

Let us see in detail what this means.

$$\langle \Gamma \rangle = \underbrace{\langle \ln p_i^{init} \rangle}_{} - \underbrace{\langle \ln p_f^{final} \rangle}_{} + \underbrace{\beta \langle E_\mu - E_m \rangle}_{}$$

$$= \sum_i p_i^{init} \ln p_i^{init} - \sum_f p_f \ln p_f + \beta \Delta Q$$

$$= - S_{init} + S_{final} + \beta \Delta Q$$

We have that $\langle \Gamma \rangle \geq 0$ implies:

$$\underbrace{\beta \, \Delta Q}_{\text{heat flow from memory into reservoir}} \geq S_{init} - S_{final}$$

For the RESET operation we have $S_{init} = \ln 2$ and $S_{final} = 0$. Thus $\Delta Q \geq + k_B T \ln 2$.

We have "derived" Landauer's principle:

"An amount of heat at least $k_B T \ln 2$ is dissipated from memory to the environment during the erasure of 1 qubit of information".

(From the first law of thermodynamics if the internal energy of the system has not changed this also entails expenditure of $\Delta W = k_B T \ln 2$ of work to erase the memory)

— THE END ☺ —