

Szilard's Theorem and Landauer's Principle

Are information entropy and thermodynamic entropy correlated?

Gerold Gründler¹

Szilard's theorem, published in 1929, and Landauer's principle, published in 1961, both postulate a relation between information and thermodynamic entropy. In this article, both theorems are described in very detail. Szilard's theorem is disproved. And Landauer's principle, though experimentally neither disproved nor confirmed, is shown to be more damaging than useful.

Contents

1. Thermodynamic entropy	2
2. Information entropy	6
3. Szilard's theorem	8
3.1 The self-measuring gas	9
3.2 A perpetuum mobile of second kind	21
3.3 The thermal memory	34
4. Landauer's principle	36
4.1. Experimental evaluations	46
4.2 The impact of thermodynamic fluctuations	53
4.3 Conclusions	56
Appendix: Bennett's account of Szilard's perpetuum mobile ..	58
References	62

¹ <mailto:gerold.gruendler@astrophys-neunhof.de>

1. Thermodynamic entropy

The first law of thermodynamics says, that it is impossible to construct a “perpetuum mobile”, i. e. a continuously running machine whose energy output per machine cycle is higher than it's energy input. For centuries, ingenious inventors like Leonardo da Vinci (1452–1519) intensively tried to construct a perpetuum mobile, but all attempts failed.

This rule, extracted from experience, can be formally stated as

the first law of thermodynamics:

$$dU = \delta Q + dW \quad (1)$$

Here U is the inner energy of the considered system, and Q and W are the heat and the work, respectively, which are fed² into the system.

It's characteristic for 19th century physics, to discern macroscopic interactions and microscopic interactions. Macroscopic interactions are mechanical work, and interactions with macroscopic electric, magnetic, and gravitational fields. Microscopic interaction is the exchange of heat.

Accordingly, the environment of the considered system is divided into a thermal bath, with which the system exchanges microscopic energy (i. e. the heat δQ), and the macroscopic environment, with which the system exchanges macroscopic energy (i. e. the work dW).

The second law of thermodynamics, which — like the first law — has been concluded from experience, is even more restrictive. It says that it is impossible to construct a “perpetuum mobile

² We apply the convention, that work or heat, which is fed from outside into the system, gets a positive sign ($W > 0, Q > 0$), while work or heat, which the system feeds towards the environment, gets a negative sign ($W < 0, Q < 0$).

of second kind". This is³ a continuously running machine which extracts per cycle the energy Q in form of heat from *one* thermal bath, and converts it into mechanical energy² $W = -Q$. Instead there must exist a second thermal bath with temperature $T_2 < T_1$, and the machine which extracts the heat Q from the heat bath at T_1 can do only the mechanical work $W = -Q \cdot (T_1 - T_2)/T_1$, and must feed² the heat $q = -QT_2/T_1$ into the bath of lower temperature.

This can be formally stated as

the second law of thermodynamics:

$$\oint \frac{\delta Q}{T} \begin{cases} = 0 & \text{if the process is reversible} \\ < 0 & \text{if the process is irreversible} \end{cases} \quad (2)$$

In an irreversible cyclic process, the system dissipates more heat to the bath than it absorbs from the bath.

Clausius [2,3] defined the entropy S due it's infinitesimal change

$$dS = \left. \frac{\delta Q}{T} \right|_{\text{rev}} = \text{entropy change of a system, which} \quad (3)$$

reversibly exchanges² heat Q with a bath of temperature T

As there is no exchange of heat with the macroscopic environment, it's entropy never changes:

$$dS_{\text{macroscopic environment}} = 0 \quad (4)$$

Using the notion of entropy, a more detailed formulation of the second law is possible, which applies not only to cyclic processes, but to arbitrary processes:

³ In the words of Kelvin [1]: "It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects."

The second law of thermodynamics:

$$dS_{\text{closed}} \stackrel{(4)}{=} dS + dS_{\text{bath}} \begin{cases} = 0 & \text{for reversible changes} \\ > 0 & \text{for irreversible changes} \end{cases} \quad (5a)$$

In words:

The entropy of a system can not decrease, unless that decrease is compensated or over-compensated by an increase of the entropy of the environment. (5b)

Don't be confused by the $>$ and $<$ signs in (2) and (5a), which result from our sign-convention².

Consequently, if the entropy of the system is reduced by $\Delta S < 0$, then the second law of thermodynamics implies that minimum the heat²

$$-\Delta Q = -T\Delta S = +T\Delta S_{\text{bath}} > 0 \quad (6)$$

must be fed from the system into the thermal environment, to increase S_{bath} by minimum $\Delta S_{\text{bath}} = -\Delta S > 0$.

Boltzmann [4] and Gibbs [5] worked out the statistical interpretation of entropy. That interpretation was based on the hypothesis that the considered system actually consists of N particles which are too small to be observed directly. According to that hypothesis, the temperature, pressure, and heat content of the system is caused by the kinetic energy of the N particles. The *microstate* of the system is its state in phase space, i.e. a point in the $6N$ -dimensional space of the positions and momenta of the N particles. It is impossible to determine the actual microstate of the system. Instead from the observed *macrostate* of the system (i.e. its volume, pressure, temperature, chemical type, etc.) it can only be concluded that the actual state in phase space must be one of M possible microstates. Boltzmann postulated that the entropy of

the system is related to the number M of the possible microstates by

$$S = k \ln M . \quad (7)$$

The probability of any of the possible microstates to be the actual microstate would be $p = 1/M$, if it would be identical for all of the possible microstates. Then Boltzmann's formula (7) could be written as

$$S = k \cdot \underbrace{M}_{1} \cdot \underbrace{p}_{M} \cdot \ln(1/p) = -kMp \ln p .$$

But in reality, for example a microstate in which most of the kinetic energy of the system is concentrated onto one single of the N particles is certainly less probable than a microstate in which the kinetic energy is distributed approximately equal to the N particles. Therefore Gibbs made this formula more flexible, allowing for different probabilities p_j of the different microstates:

$$S = -k \sum_{j=1}^M p_j \ln p_j \quad \text{with} \quad \sum_{j=1}^M p_j = 1 \quad (8)$$

This is the entropy of a system, which assumes with probabilities p_j various microstates j , all of which are compatible with the observed (macro)state of that system.

Actually there exists a basic, very important difference in the notions of entropy as defined by Clausius and Gibbs. In [6], that difference, and it's far-reaching consequences, are explicated. For our present purpose — the discussion of Szilard's theorem and Landauer's principle —, however, that difference can be ignored. But the difference inbetween thermodynamic entropy (be it in the definition of Clausius or Gibbs) on the one hand, and information entropy on the other hand, will turn out to be essential in our below considerations.

2. Information entropy

Consider a variable X , which can assume the n different discrete values $\{x_1, \dots, x_n\}$. X may for example be the result of throwing a die (then $n = 6$ for a conventional die), or the result of tossing a coin (then $n = 2$), or the signal received from a telecommunication line.

We now want to define the *information content* of a particular value x_j of the variable X . The information content of a result with a fair die is $3\times$ as high as the information content of a result with a fair coin, because the result of tossing the coin can be guessed in advance (before the coin is actually thrown and the result observed) with probability $1/2$, while the result of throwing the die can be guessed in advance only with probability $1/6$. Obviously the information content $I(x_j)$ should be a function of P_j^{-1} , with P_j being the probability of $X = x_j$:

$$I(x_j) = \text{function}\left(\frac{1}{P_j}\right) \quad (9a)$$

If two observations of the variable X are independent (for example the result x_a of throwing a die, and the result x_b of a later throw with the same die), then the information content of the combined results is additive:

$$I(x_a \cdot x_b) = I(x_a) + I(x_b) \quad (9b)$$

Thus $I(x_j)$ must be a logarithmic function of x_j . In his “Mathematical Theory of Communication” [7], published in 1948, Claude Elwood Shannon (1916–2001) chose the logarithm to the basis 2. Considering (9a) and (9b), *information content* may be defined as follows:

$$I(x_j) = \log_2\left(\frac{1}{P_j}\right) = -\log_2 P_j \quad (10)$$

Furthermore Shannon [7] defined the *information entropy* of the variable X as the mean value of the information contents $I(x_j)$ of all values x_j which the variable X does assume with probabilities P_j :

$$H(X) = \sum_{j=1}^n P_j I(x_j) \stackrel{(10)}{=} - \sum_{j=1}^n P_j \log_2 P_j \quad \text{with} \quad \sum_{j=1}^n P_j = 1 \quad (11)$$

Here H is the Greek capital letter Eta. Famously it was John v. Neumann, who in a discussion with Shannon suggested to name the quantity (11) entropy, “because anyway nobody knows what entropy is.” Indeed the similarity of the formulas for Gibb’s thermodynamic entropy (8) and Shannon’s information entropy (11) is striking.⁴

Note, however, that I took care to discern the probabilities p_j of unobservable microstates in phase space, all of which are compatible with the observed macrostate of the considered system, from the probabilities P_j of observable macrostates, which furthermore may be completely different from states in phase space, due to lower and upper case characters. It is controversial amongst physicists, whether the probabilities p_j and P_j really are of basically different nature and consequently must be kept separated, or whether the difference is irrelevant, and the p_j and P_j — and consequently thermodynamic entropy and information entropy — should be considered essentially equivalent. See the discussion in section 4 below.

⁴ In this article, the word “information” is always explicitly stated in case of information entropy. When the word “entropy” is used without an explicit specifier “information” or “thermodynamic”, then always thermodynamic entropy is meant.

3. Szilard's theorem

In 1929, Leo Szilard (1898–1964) published an article [8], in which he tried to demonstrate a deep relation between the second law of thermodynamics and the accumulation of information. His assertion may appropriately be called

Szilard's theorem:

Any measurement and storage of the result causes per bit of measurement result minimum the thermodynamic entropy decrease (12a)

$$\Delta S = -k \ln 2 < 0$$

of measured object & measurement instrument & memory.

Szilard did not explicitly formulate this theorem, but from the context of his article there can be no doubt that this is exactly what he wanted to point out.

According to the second law of thermodynamics, an entropy reduction of minimum $\Delta S = -k \ln 2$ of measured object & measurement instrument & memory must be compensated by an entropy increase of minimum

$$\Delta S_{\text{bath}} = +k \ln 2 > 0 \quad (12b)$$

of the environment, i. e. at temperature T minimum the heat²

$$-Q = kT \ln 2 \quad (12c)$$

must be dissipated to the thermal bath. According to the first law, the energy (12c) must come from the work

$$W = kT \ln 2, \quad (12d)$$

which minimum needs to be done per bit of measurement result, to accomplish the measurement.

In his article, Szilard discussed the entropy decrease (12a), but he was silent about the related work (12d) and heat transfer (12c). Thus in his examples, to be discussed below, it is unclear how and when exactly that work and heat transfer are to happen.⁵

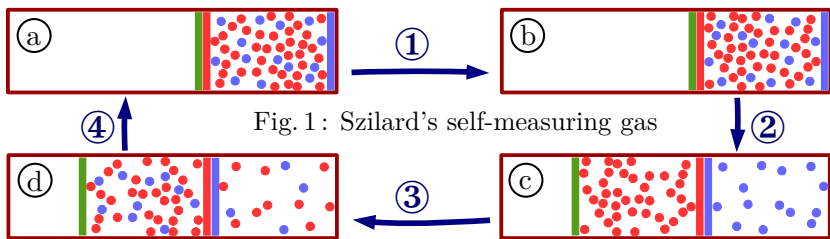
To motivate and explicate his idea, Szilard considered three examples: A self-measuring gas, a perpetuum mobile of second kind, and a measuring instrument with thermal memory. A general theorem like (12a) can of course not be proved by the presentation of some few examples. Indeed we will see that the self-measuring gas example actually rests on a confusion of information entropy and thermodynamic entropy,⁶ that the perpetuum mobile on closer scrutiny turns to an argument against (12a), and that the thermal memory can easily be modified to an argument against (12a). In the sequel, Szilard's three example machines are presented and discussed one by one.

3.1. The self-measuring gas

To demonstrate a relation between information and thermodynamic entropy, Szilard considered a gas of atoms, which have — besides their three translational degrees of freedom — an internal degree of freedom, which can assume the two discrete values X^+ or X^- . This property is changing stochastically, such that an atom with X^+ changes at an unpredictable point of time to X^- , and at a later unpredictable point of time back to X^+ , and so on. While it is impossible to predict the point of time, at which a particular atom changes the X property, the time constant of the property change for a statistically huge number of atoms can of course be

⁵ A modification of Szilard's idea due to Bennett, in which this question is precisely answered, will be described in the [appendix](#).

⁶ As Szilard published his article 20 years before Shannon invented information entropy, this confusion is not really surprising.



precisely determined. We will use the Greek character τ for the time constant of a full cycle from X^+ to X^- and back to X^+ .

The probability to observe an atom at some point of time with property X^+ is P^+ , and the probability to observe it with property X^- is P^- , with $P^+ + P^- = 1$.

Attached to each atom is a tiny measurement instrument and a 1-bit memory. Upon command of an observer, the instrument measures whether the atom is in state X^+ or X^- . If the result is X^+ , the memory is set to Y^+ . If the result is X^- , the memory is set to Y^- .

Szilard assumed to have a huge number N of these “molecules” available, each molecule consisting of one atom, one measuring instrument, and one memory. The molecules form an ideal gas, which is enclosed in a vessel, as sketched in figure 1. Red dots symbolize atoms which at this point of time are in state X^+ , blue dots symbolize atoms which are in state X^- . While “red” molecules are permanently changing to “blue”, and “blue” molecules are changing to “red”, the fraction N^+/N of “red” molecules deviates only negligibly from P^+ , and the fraction N^-/N of “blue” molecules deviates only negligibly from P^- at any time, because N is a huge number.

The gas is enclosed between movable partitions, whose positions are steered from outside. The blue partition right of the gas is permeable for molecules in state X^- , i. e. the “blue” molecules, but

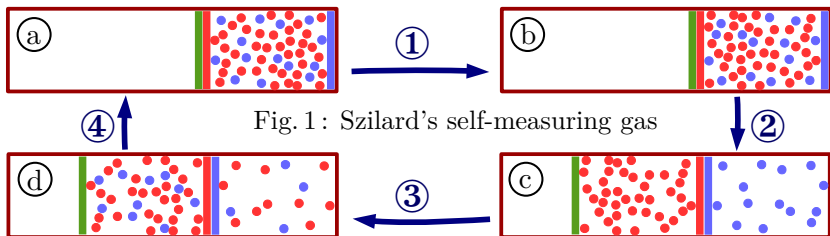


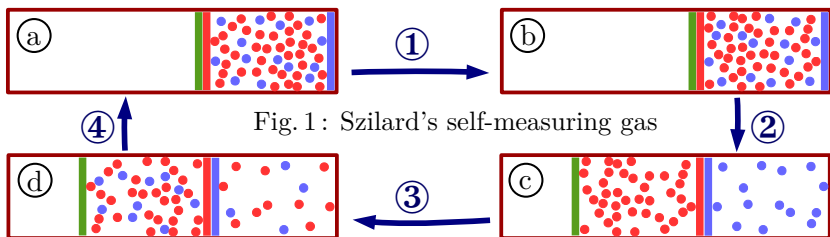
Fig. 1: Szilard's self-measuring gas

impermeable for molecules in state X^+ , i. e. the “red” molecules. The red partition left of the gas is permeable for molecules in state X^+ , but impermeable for molecules in state X^- . The movable green partition is impermeable for all molecules. The volume left of the green partition is evacuated. The vessel is embedded into a thermal bath of temperature T .

Now Szilard considered this cyclic process:

- Ⓐ Long ago, very many time constants τ before the start of the cycle, the outside observer has ordered a measurement. Upon this order, each measurement instrument has measured the state of it's atom (X^+ or X^-), and set it's memory accordingly to Y^+ or Y^- . Thus immediately after that measurement, $N^+ = P^+N$ molecules have been in state X^+ and their memories were set to Y^+ , while $N^- = P^-N$ molecules have been in state X^- and their memories were set to Y^- . The atom state X and the memory state Y of each molecule had been perfectly correlated at that point of time.

Since then, almost all atoms have many times switched back and forth between X^+ and X^- , while the Y values of the memories have not changed since the measurement. Thus there are now still $N^+ = P^+N$ molecules in state X^+ and $N^- = P^-N$ molecules in state X^- , and N^+ memories in state Y^+ and N^- memories in state Y^- . But the correlation between the X and the Y of each molecule is completely lost.



This is the state ①, from which the cyclic process starts.

- ① In the first process step, the outside observer orders a new measurement. This measurement changes the information entropy of the gas. To see that, consider the probabilities, with which an arbitrarily picked single molecule might have the properties $\{X, Y\}$ before and after the measurement:

properties	probability before measurement	probability after measurement
$\{X^+, Y^+\}$	$P^+ P^+$	P^+
$\{X^+, Y^-\}$	$P^+ P^-$	0
$\{X^-, Y^+\}$	$P^- P^+$	0
$\{X^-, Y^-\}$	$P^- P^-$	P^-

Before the measurement, the molecules could be in four different $\{X, Y\}$ states, while immediately after the measurement they can be in only two different $\{X, Y\}$ states. According to Shannon's formula

$$H(X) \stackrel{(11)}{=} - \sum_{j=1}^n P_j \log_2 P_j$$

for information entropy (which was yet unknown in 1929), this is the information entropy reduction of the gas in course of the measurement:

$$\begin{aligned}
 \Delta H &= N \left(-P^+ \log_2 P^+ - P^- \log_2 P^- + \right. \\
 &\quad + P^+ P^+ \log_2(P^+ P^+) + P^+ P^- \log_2(P^+ P^-) + \\
 &\quad \left. + P^- P^+ \log_2(P^- P^+) + P^- P^- \log_2(P^- P^-) \right) = \\
 &= N \left(-P^+ \log_2 P^+ - P^- \log_2 P^- + \right. \\
 &\quad + 2P^+(P^+ \log_2 P^+ + P^- \log_2 P^-) + \\
 &\quad \left. + 2P^-(P^+ \log_2 P^+ + P^- \log_2 P^-) \right) = \\
 &= N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) < 0 \tag{13}
 \end{aligned}$$

Here $P^+ + P^- = 1$ has been used. Szilard, however, assumed — using Gibbs’ entropy formula (8) — this *thermodynamic* entropy reduction of the gas:

$$\Delta S \stackrel{(13)}{=} Nk \left(P^+ \ln P^+ + P^- \ln P^- \right) < 0 \tag{14}$$

This is a remarkable extension of the entropy notion as used by Clausius, Boltzmann, and Gibbs, for two reasons: First, in Gibbs’ formula (8) the p_j are the probabilities of *inaccessible* microstates, which all are compatible with the observed macrostate of the considered system, while the P_j in Shannon’s formula (11) are the probabilities of a macroscopically accessible property of the considered system — in this case the $\{X, Y\}$ property of Szilard’s “molecules”. Second, the inaccessible microstates of Boltzmann and Gibbs are dynamic states in phase space, which are to explain the heat content and temperature of the considered system. The property $\{X, Y\}$, however, is not at all related to heat content and temperature of Szilard’s artificial “gas”: The ratio $N^+/N^- = P^+/P^-$ would not be affected by a change of the temperature T of the system, while a change of temperature would of course change

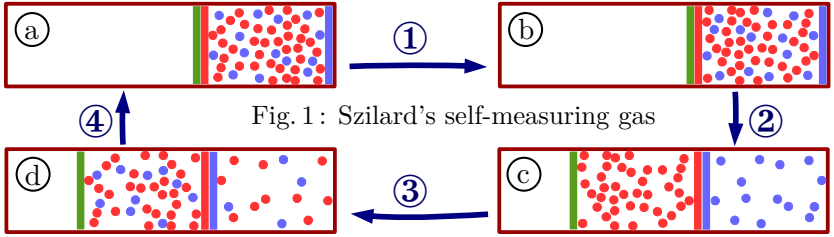


Fig. 1: Szilard's self-measuring gas

the phase space accessible to the gas, and consequently the probabilities p_j in Gibbs' formula (8).

- ② Immediately after the measurement the green partition (which is impermeable for all molecules) is moved left, and synchronously the blue partition (which is permeable for blue molecules, but impermeable for red molecules) is moved left at same speed. Thereby the sub-ensemble of red molecules is moved left with unchanged volume V_0 , while the sub-ensemble of blue molecules stays with unchanged volume V_0 at it's previous position. The movement is stopped when the blue partition touches the red partition. No work needs to be done to accomplish the shift, nor does the shift induce any change of temperature or thermodynamic entropy.⁷
- ③ If the shift is completed within a small time interval $\Delta t \ll \tau$ after the measurement, then immediately after the shift there are $N^+ = P^+N$ molecules in the left volume, which (almost) all are in the "red" state X^+ and have memory content Y^+ .

⁷ To see this clearly, imagine that — without changing the volume V_0 — the red gas is by means of a semipermeable partition compressed into the left half of this volume, and the blue gas is by means of a semipermeable partition compressed into the right half of this volume, such that we get N^+ red molecules and no blue molecules compressed into the volume $V_0/2$ left of the semipermeable partitions, and N^- blue molecules and no red molecules compressed into the volume $V_0/2$ right of the semipermeable partitions. To achieve this reversible demixing, the work

And there are $N^- = P^- N$ molecules in the right volume, which (almost) all are in the “blue” state X^- and have memory content Y^- .

	state (b)	state (c)	
		left volume	right volume
$\{X^+, Y^+\}$	P^+	1	0
$\{X^-, Y^+\}$	0	0	0
$\{X^+, Y^-\}$	0	0	0
$\{X^-, Y^-\}$	P^-	0	1

Thus in step (2) a decrease of information entropy has happened:

$$\begin{aligned} \Delta H = & \underbrace{-N^+ \left(\overbrace{1 \log_2 1}^0 + 3 \cdot \overbrace{0 \log_2 0}^0 \right)}_{\text{left volume}} - \\ & \underbrace{-N^- \left(3 \cdot 0 \log_2 0 + 1 \log_2 1 \right)}_{\text{right volume}} + \\ & + N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) = \end{aligned}$$

$$\Delta W \stackrel{(23)}{=} kTN^+ \int_{V_0}^{V_0/2} \frac{dV}{V} + kTN^- \int_{V_0}^{V_0/2} \frac{dV}{V} \quad (15a)$$

must be done. After the demixing, each gas is reversibly expanded to volume V_0 , i. e. to state (c). Thereby the gas does the work

$$-\Delta W \stackrel{(23)}{=} kTN^+ \int_{V_0/2}^{V_0} \frac{dV}{V} + kTN^- \int_{V_0/2}^{V_0} \frac{dV}{V} = -(15a). \quad (15b)$$

Thus the overall net exchange of energy between gas and environment is zero, and there is indeed no change of thermodynamic entropy in process step (2).

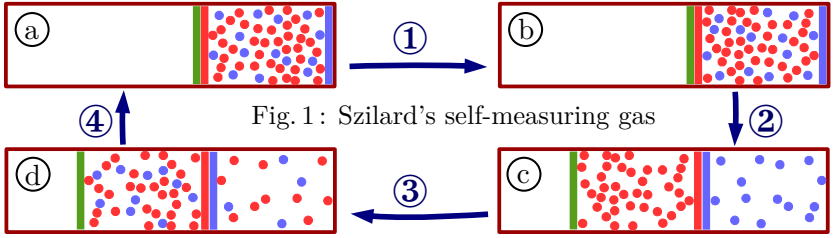


Fig. 1: Szilard's self-measuring gas

$$= N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) = \Delta H < 0 \quad (16)$$

But, while Szilard in step ① had concluded from the information entropy reduction (13) the thermodynamic entropy reduction (14), now in step ② Szilard stuck to the conventional definition of thermodynamic entropy, i. e. he assumed no change of thermodynamic entropy, despite the change (16) of information entropy.

- ③ The third process step is merely a waiting time interval $\Delta t \gg \tau$. During this time, the atoms in both volumes change back and forth between X^+ and X^- , while their memory contents (i. e. the Y -values) don't change. Thereby the probabilities that a blindly picked molecule will be in state X^+ or X^- (i. e. the information entropy of the X property) changes as displayed in this table:

properties	left volume probability in		right volume probability in	
	state ③	state ④	state ③	state ④
$\{X^+, Y^+\}$	1	P^+	0	0
$\{X^-, Y^+\}$	0	P^-	0	0
$\{X^+, Y^-\}$	0	0	0	P^+
$\{X^-, Y^-\}$	0	0	1	P^-

Consequently this is the change of information entropy:

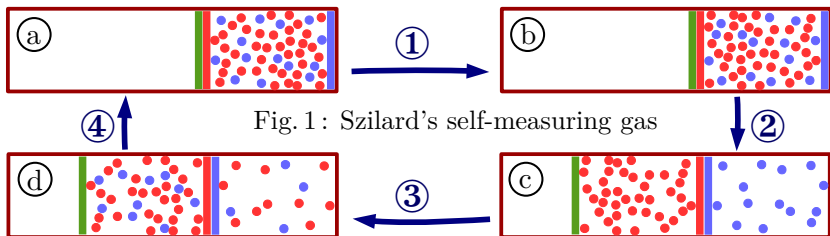


Fig. 1: Szilard's self-measuring gas

$$\begin{aligned}
 \Delta H &= N^+ \left(-P^+ \log_2 P^+ - P^- \log_2 P^- + \overbrace{1 \log_2 1}^0 + \overbrace{0 \log_2 0}^0 \right) + \\
 &\quad \underbrace{\hspace{15em}}_{\text{left volume}} \\
 &\quad + N^- \left(-P^+ \log_2 P^+ - P^- \log_2 P^- + 0 \log_2 0 + 1 \log_2 1 \right) = \\
 &\quad \underbrace{\hspace{15em}}_{\text{right volume}} \\
 &= -N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) > 0 \tag{17}
 \end{aligned}$$

This time Szilard again identified the change of information entropy as a change of thermodynamic entropy, and assumed by means of Gibbs' entropy formula (8) this thermodynamic entropy change during process step ③:

$$\Delta S \stackrel{(17)}{=} -Nk \left(P^+ \ln P^+ + P^- \ln P^- \right) > 0 \tag{18}$$

- ④ At start of the last process step, a further semipermeable partition, which is permeable for molecules with memory content Y^- , but impermeable for molecules with memory content Y^+ , is vertically inserted directly right of the two semipermeable partitions. And directly left of the semipermeable partitions a further semipermeable partition is vertically inserted, which is permeable for molecules with memory content Y^+ , but impermeable for molecules with memory content Y^- . Then the

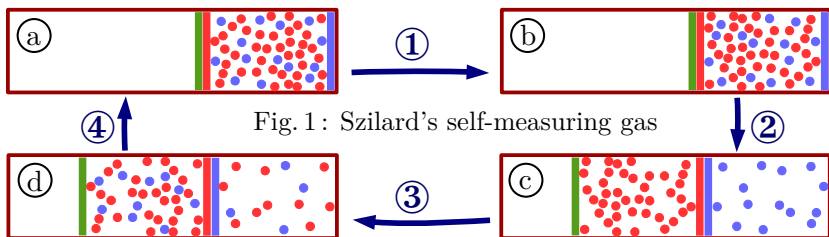


Fig. 1: Szilard's self-measuring gas

two partitions, which are semipermeable for X^+ resp. X^- are vertically pulled out. The change of partitions doesn't need energy, and does not change any parameter of the system.

Remember that *all* molecules right of the semipermeable partitions have memory content Y^- , and *all* molecules left of the semipermeable partitions have memory content Y^+ . Now the partition, which blocks Y^+ but not Y^- , and the very left (green) partition, which blocks all molecules, are shifted right at same speed, until they touch the right vessel wall respectively the third partition. Thereby the Y^+ gas is shifted towards right at unchanged volume V_0 , and mixed into the Y^- gas, which as well keeps it's unchanged volume V_0 .

As both the Y^+ gas and the Y^- gas are ideal gases, and their volumes V_0 are constant during the mixing, no work is done during the mixing, there is no heat exchange with the bath, and no change of thermodynamic entropy. (That could be double-checked with a quite similar computation as presented in footnote ⁷ for process step ②.)

But of course there is a change of information entropy: Before the shift, the probability to pick a Y^+ molecule in the left compartment was 1, and the probability to pick a Y^+ molecule in the right compartment was 0, while the probability to pick a Y^- molecule in the left compartment was 0, and the probability to pick a Y^- molecule in the right compartment

was 1. After the shift, the probability to get by a blind pick a Y^+ molecule is P^+ , and the probability to pick a X^- molecule is P^- :

	state ④		state ③
	left volume	right volume	
$\{X^+, Y^+\}$	P^+	0	P^+P^+
$\{X^-, Y^+\}$	P^-	0	P^-P^+
$\{X^+, Y^-\}$	0	P^+	P^+P^-
$\{X^-, Y^-\}$	0	P^-	P^-P^-

Thus the change of information entropy in step ④ is

$$\begin{aligned}
 \Delta H &= N \left(-P^+P^+ \log_2(P^+P^+) - P^+P^- \log_2(P^+P^-) - \right. \\
 &\quad \left. - P^-P^+ \log_2(P^-P^+) - P^-P^- \log_2(P^-P^-) + \right. \\
 &\quad \left. + P^+ \log_2 P^+ + P^- \log_2 P^- \right) \stackrel{(13)}{=} \\
 &= N \left(-P^+ \log_2 P^+ - P^- \log_2 P^- \right) > 0 \quad (19)
 \end{aligned}$$

In this step, however, Szilard again stucked to the conventional definition of thermodynamic entropy, and ignored the change of information entropy.

At the end of process step ④, state ③ is reached again, and the process cycle is closed.

In total, we observed these changes of information entropy per cycle of Szilard's self-measuring gas system:

$$\text{step ①: } \Delta H \stackrel{(13)}{=} +N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) \quad (20a)$$

$$\text{step ②: } \Delta H \stackrel{(16)}{=} +N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) \quad (20b)$$

$$\text{step ③: } \Delta H \stackrel{(17)}{=} -N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) \quad (20c)$$

$$\text{step } \textcircled{4} : \quad \Delta H \stackrel{(19)}{=} -N \left(P^+ \log_2 P^+ + P^- \log_2 P^- \right) \quad (20d)$$

Note that $\Delta H = (20)$ sums up to zero. The machine cycle indeed is a true cycle: At the end, we have exactly the same amount of information about the $\{X, Y\}$ property available as at the begin of the cycle.

We noted that there is *no* thermodynamic entropy change according to the definition of thermodynamic entropy by Clausius, Boltzmann, and Gibbs in any step of the cycle. Szilard, however, assumed these changes of thermodynamic entropy:

$$\text{step } \textcircled{1} : \quad \Delta S \stackrel{(13)}{=} +Nk \left(P^+ \ln P^+ + P^- \ln P^- \right) \quad (21a)$$

$$\text{step } \textcircled{3} : \quad \Delta S \stackrel{(17)}{=} -Nk \left(P^+ \ln P^+ + P^- \ln P^- \right) \quad (21b)$$

As Szilard — in accord with the conventional definition of thermodynamic entropy — assumed no thermodynamic entropy changes in steps $\textcircled{2}$ and $\textcircled{4}$ of this cyclic process, and as the alleged thermodynamic entropy changes of steps $\textcircled{1}$ and $\textcircled{3}$ exactly cancel, Szilard argued: If we would have overlooked the thermodynamic entropy reduction due to the measurement in step $\textcircled{1}$, then $\oint dS$ would be different from zero, even though the process is clearly cyclic. Consequently the self-measuring gas example was in his opinion a support for his theorem (12a).

Szilard's argument is clearly circular, because he *two* times confused information entropy and thermodynamic entropy. If he had correctly inserted zero thermodynamic entropy change in step $\textcircled{3}$, then his argument for a thermodynamic entropy change in step $\textcircled{1}$ would completely vanish.

Szilard explained that it seemed unlikely to him that the entropy decrease due to a measurement should depend on the particular values of P^+ and P^- . Considering that for arbitrary $P^+ = 1 - P^-$

$$-P^+ \ln P^+ - P^- \ln P^- \leq -\frac{1}{2} \ln \frac{1}{2} - \frac{1}{2} \ln \frac{1}{2}, \quad (22)$$

he decided to insert $\ln 2$ into (12a).

But why couldn't we *define*: Entropy is entropy, no matter whether it has traditionally been called thermodynamic entropy or information entropy? I.e. why shouldn't we state *by definition* that Gibbs' probabilities p_j of different microstates and Shannon's probabilities P_j of different macrostates are equivalent and may be inserted interchangeably, as Szilard has done two times? In section 4.3 good reasons will be presented, why these two types of entropy should better be kept strictly separated.

3.2. A perpetuum mobile of second kind

Szilard constructed this machine such, that an entropy reduction due to measurement is needed, to cure an alleged violation of the second law of thermodynamics. We will see, however, that actually that measurement is *not* correlated with any change of thermodynamic entropy.

The perpetuum mobile of second kind is sketched in fig. 2 on the next page in a slightly improved form as suggested by Bennett [9,10]. If this machine would be functional (of course it is not), it would work like this:

- Ⓐ The machine cycle starts from state Ⓐ. One single rare-gas atom is trapped in a vessel of volume V_m . The vessel is embedded into a thermal bath with temperature T . Movable pistons can be shifted from left and right into the vessel. The atom exerts a pressure $p_{(a)}$ onto the pistons and onto the walls of the vessel, which can be computed from the equation of an ideal gas, consisting of 1 atom:

$$p_{(a)} = \frac{kT}{V_m} \quad (23)$$

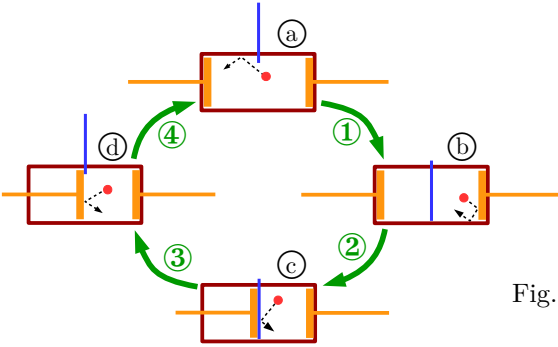


Fig. 2: Szilard's perpetual mobile of second kind

- ① In the first process step, a partition is inserted, which divides the vessel into 2 cavities of size $V_m/2$ each. It's of no relevance whether the atom is trapped by chance in the left or in the right cavity. No (or only negligible) energy is required to shift-in the partition.⁸ As the volume of the ideal gas (consisting of the 1 rare gas atom) is changed without doing mechanical work on the gas or extracting work from the gas, the atom's kinetic energy, and hence the temperature of the gas, is not changed. But there is a change of the volume, which is accessible to the 1-atom-gas, and the pressure, which it exerts onto the vessel walls, the partition, and the pistons:

$$p_{(b)} = \begin{cases} kT/(V_m/2) = 2p_{(a)} & \text{in the cell with the atom} \\ 0 & \text{in the empty cell} \end{cases} \quad (24)$$

Thus the mean value over both cells is $\overline{p_{(b)}} = p_{(a)}$.

As in state (b) only half as many microstates in phase-space are accessible to the atom as in state (a), the thermodynamic

⁸ In this analysis, we assume van der Waals forces to be negligible. Later we will argue that this is a misleading over-simplification.

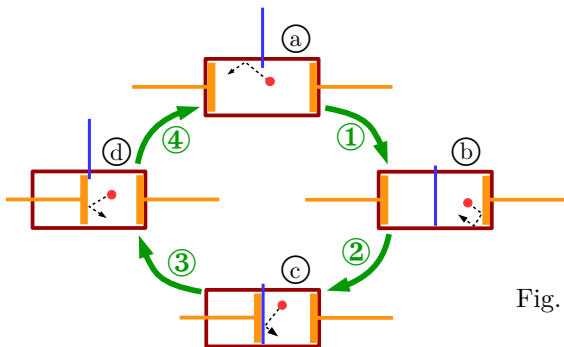


Fig. 2: Szilard's perpetual mobile of second kind

entropy of the 1-atom gas has changed in process step ① by

$$\Delta S = -k \ln 2 < 0 . \quad (25)$$

This entropy decrease is not compensated by an entropy increase of the bath, clearly violating the second law (5). This strange fact makes Szilard's machine a perpetual mobile of second kind.

② In the second process step, a small force

$$f \ll p_{(b)} A ,$$

with A being the area of the pistons, is exerted from outside onto *both* pistons, trying to shift them into the vessel. One of the pistons does not move because the small force f can not overcome the pressure $p_{(b)}$, and consequently no work is done. The other piston is moved inside, until it touches the partition. No (or only negligible) work is done with this piston, because there is no (or only negligible) resistance against it's inside move. And there is no change of thermodynamic entropy in this process step.

③ In the third process step, appropriate mechanisms are attached

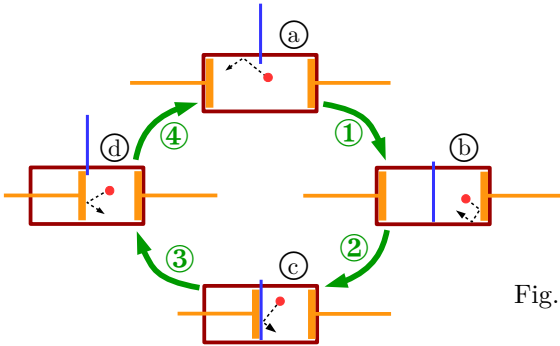


Fig. 2: Szilard's perpetual mobile of second kind

to *both* pistons⁹ outside the vessel, which will lift loads against the earth's gravitational field, provided the pistons should be pushed towards the outside. Then the partition is pulled-out. This requires no (or only negligible) energy, and there is no change of entropy in this process step.

- ④ In the last process step, one of the pistons is moved due to the pressure exerted by the atom, and lifts the load of mass M up to height h . The useful mechanical work done by the machine is²

$$-W = Mgh = \int_{V_m/2}^{V_m} dV p(V) \stackrel{(23)}{=} kT \int_{V_m/2}^{V_m} \frac{dV}{V} = kT \ln 2 ,$$

⁹ To extract work from Szilard's original machine [8], the user needs to know upfront in which of the two compartments the atom is trapped. That's of course a faulty design. Bennett, though he suggested [9, 10] the improved form of Szilard's machine, still believed that an upfront measurement of the atom's position would be needed to extract energy from the machine, and — most important for Bennett's interpretation of this process — that the result of that measurement must be stored in a memory, and later be erased from that memory, allegedly at the expense of energy. I will discuss Bennett's point of view in the [appendix](#).

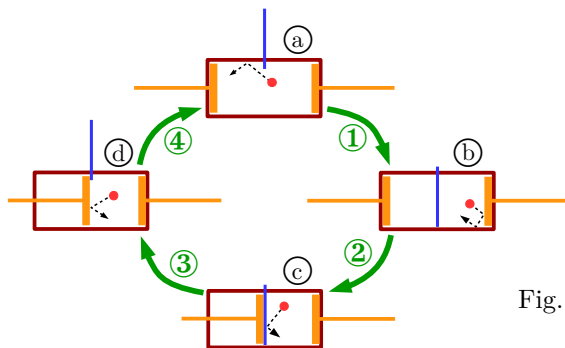


Fig. 2: Szilard's perpetuum mobile of second kind

with g being the gravitational acceleration. This energy comes from the heat

$$Q = -W = kT \ln 2 \quad (26)$$

which the atom absorbs from the thermal bath during process step ④.

As the phase space, which is accessible to the atom, is twice as large in state ① than in state ②, the entropy of the 1-atom gas increases in process step ④ by

$$\Delta S = +k \ln 2 > 0 , \quad (27a)$$

and the entropy change of the bath is

$$\Delta S_{\text{bath}} = -(27a) = -k \ln 2 < 0 . \quad (27b)$$

Note that $Q = W = 0$ in each of the process steps ①, ②, ③. Only in process step ④ the heat transfer $Q = -W > 0$ happens. Furthermore the entropy increase (27a) exactly compensates the entropy decrease (25), while the entropy decrease (27b) of the bath is not compensated.

Thus the machine does exactly, what is impossible according to the second law: It is a cyclic working machine, which in each cycle absorbs heat $Q > 0$ from a bath and converts it into usable work $-W = Q > 0$, *even though there exists no temperature difference between the bath, the machine, and the rest of the environment*. Stated in terms of entropy: This machine continuously decreases the entropy of the closed system gas & bath by $\Delta S_{\text{gas \& bath}} = -k \ln 2$ per cycle.

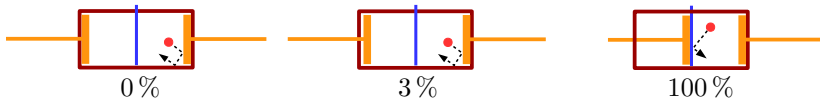
Here Szilard saw an argument for his theorem (12a), which assigns an entropy reduction to *any* measurement. There is indeed a measurement in the cyclic process fig. 2: After the partition has been inserted in process step ①, we do not yet know whether the atom has been trapped in the left or in the right cavity.

We get that information due to process step ②, which clearly is a measurement and storage of the result. When a small force is applied to both pistons, but only one of them yields and moves in, then indeed the position of the atom has been measured, with the result that it must be in the other cavity. Of course we do not need⁹ this knowledge to successfully drive the perpetuum mobile and extract useful work from it. But the measurement has anyway been done, and the result (left piston in or right piston in) has objectively come into existence, and is stored by the moved-in position of the piston. It's of no relevance whether anybody looks and notices the result.

This measurement merely implies a decrease of *information* entropy by 1 bit. According to Szilard's theorem (12a), however, there must be a decrease $\Delta S = -k \ln 2$ of *thermodynamic* entropy due to this measurement and result storage. Consequently, if Szilard's theorem should be correct, then — due to the second law of thermodynamics — the work $W = kT \ln 2$ must be done in course of the measurement, and thereby the heat² $-Q = kT \ln 2$ be dissipated into the bath. This work just compensates the work

done by the machine in process step ④, and the related heat extraction from the bath. Therefore the machine displayed in fig. 2 is no perpetuum mobile of second kind, and the second law of thermodynamics is “saved”, if Szilard’s theorem (12a) should be correct.

A severe problem of Szilard’s theorem becomes visible, however, when process step ② is analyzed in detail. The three sketches show Szilard’s machine after



of this process step have been completed. Let’s assume that we can clearly discern which piston moves, and which not, as soon as one piston has moved 1 % of it’s full stroke. We could stop the process at this moment, and still would have 100 % of the measurement result, i. e. we would know for sure whether the atom is trapped in the left or in the right cavity. Thus the measurement proper (including storage of the result!) is completed after 1 % of the piston stroke, no additional information can be gathered during the remaining 99 % of process step ②. Consequently, if we had a sufficiently sensitive calorimeter at our disposal, we would have registered 100 % of the heat $kT \ln 2$, which is dissipated according to Szilard’s theorem due to the measurement, within the first 1 % of this process step. No further heat would be dissipated during the remaining 99 % of the piston stroke.

The first law of thermodynamics must be preserved. This means that the dissipated heat does not come for free out of nothing. Instead — different from our previous assumption that the piston of the empty cavity can be moved with doing almost no work — the work $W_{\text{empty}} = kT \ln 2$ must be done to move this piston. And all of this work must be done during the measurement, i. e. within the first percent of the full piston stroke. With $L_0 = (V_m/2)/A$

being the full length of the piston stroke, and A being the area of the piston, the work

$$W_{\text{empty}} = \int_0^{1\% \cdot L_0} dL F_{\text{empty}} = 1\% \cdot L_0 \cdot \bar{F}_{\text{empty}} = kT \ln 2 \quad (28a)$$

$$\bar{F}_{\text{empty}} = \frac{100}{L_0} kT \ln 2 \quad (28b)$$

must be done, with F being the force resisting the inward move of the piston, and \bar{F} being the mean value of F during the first percent of the stroke. But as soon as the measurement result is available, F shrinks to zero, and the remaining 99% of the stroke can be done with no (or only negligible) work. Lets compare this work with the work that must be done to move the piston of the other cavity (the cavity with the atom) 1% in:²

$$W_{\text{atom}} = -kT \int_{V_m/2}^{99\% \cdot V_m/2} \frac{dV}{V} = -kT \ln 0.99 \quad (28c)$$

$$\bar{F}_{\text{atom}} = -\frac{100}{L_0} kT \ln 0.99 \stackrel{(28b)}{\approx} \frac{\bar{F}_{\text{empty}}}{69} \quad (28d)$$

\bar{F}_{empty} is almost 69 times as large as \bar{F}_{atom} ! Consequently, when the force applied to both pistons is slowly increased, then — contrary to our previous assumption — the piston of the cavity with the atom will yield first, but not the piston of the empty cavity. But ... stop! Once we have learned which piston yields first, the 1% move of this piston gives us the full measurement result. According to Szilard's theorem, we can not get the measurement result cheaper than by the work $kT \ln 2$. Consequently neither \bar{F}_{atom} nor \bar{F}_{empty} can be smaller than $100 kT \ln 2 / L_0$.

This situation is strange, to say the least. And it becomes even stranger, if we consider this scenario: Imagine that due to a leak

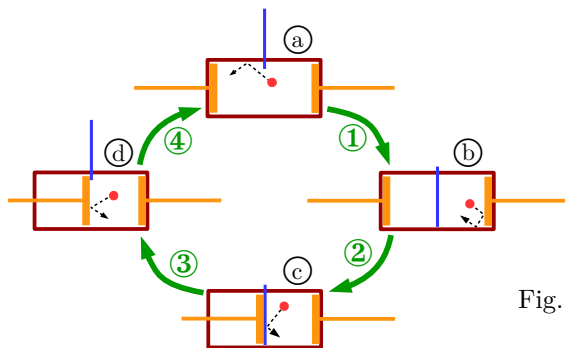


Fig. 2: Szilard's perpetual mobile of second kind

the atom unfortunately has suddenly disappeared, and the vessel is empty in process state (a). Now we perform process step (2). Both pistons will move in and touch the partition. From this result we get the information about the disaster. This information clearly is a measurement result. Consequently we will not get this information for free, but must spend minimum the work $kT \ln 2$. Or two times that work, because this is a 2-bit-information?

At first sight, Szilard's theorem seemed to cure the perpetual mobile of second kind. But once we dig into the details, the situation becomes more and more absurd. Given this situation, it seems advisable to look a little bit more skeptical onto Szilard's alleged perpetual mobile of second kind.

A valid "gedankenexperiment" may assume a setup which is today (and may be forever) beyond our technical capabilities and available resources. But it must not be in conflict with well-established laws of nature. Szilard's perpetual mobile fig. 2 does *not* meet this requirement. Szilard's essential error was, that he neglected quantum fluctuations, which induce van der Waals forces. It's easy to see that van der Waals forces actually are dominating the machine, while the tiny pressure due to the 1-atom-gas is almost negligible. Let's explicitly compute the contribution of these two

types of forces in process step ④:

With V_m being the total volume (both compartments) of the vessel, A being the area of one piston, and $V = V_m - L \cdot A$ being the variable volume of the 1-atom gas during the stroke, the gas exerts onto the piston the force

$$F(L) = p \cdot A \stackrel{(23)}{=} \frac{kTA}{V} = \frac{kTA}{V_m - LA} . \quad (29)$$

To get a strong force, the temperature must be as high as possible, and the volume as small as possible. Let's assume $T = 1000$ K, $V_m/2 = 1 \mu\text{m}^3$, and $A = 1 \mu\text{m}^2$.

The attractive van der Waals force between two metal plates of area A at distance L is approximately [11, eq. (36)]

$$F(L) = \frac{\pi^2 \hbar c A}{240 L^4} . \quad (30)$$

The sum of the van der Waals forces, which the right piston and the left vessel wall exert onto the left piston in course of process

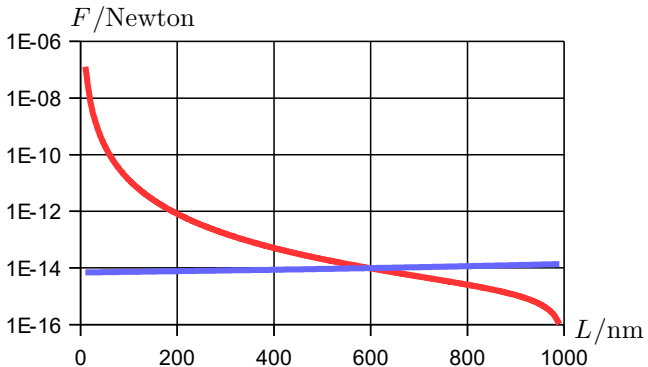


Fig. 3: Force due to pressure of the 1-atom-gas (blue line) and van der Waals force (red line)

step ④ is indicated in fig. 3 as a red line, while the force due to the pressure of the 1-atom-gas is indicated by the blue line. Note the logarithmic scale! The effect of van der Waals forces is by many orders of magnitude stronger than the gas pressure, which shall drive Szilard's machine.

To avoid this effect, we could construct the vessel as a very long tube, such that the pistons never come close to the left or right vessel walls. But that would only shift the problem, because then we would get a huge net effect due to van der Waals attraction between the piston and the partition in process step ②. No matter how we construct the details of Szilard's machine, van der Waals forces definitively are not negligible. Just the opposite: Van der Waals forces are dominating, while the tiny pressure due to the single atom is almost negligible.

This fact does not yet disprove Szilard's argument. *If* the van der Waals forces should be 100 % conservative, then we would in course of a full machine cycle get back 100 % the work, which needs to be done in some steps of the cycle against the van der Waals forces.

But that is not to be expected, by the very nature of van der Waals forces. Van der Waals forces are caused by induced dipole moments of atoms and molecules, which again are caused by spontaneously fluctuating dipole moments of other atoms and molecules [12]. Due to the fluctuating van der Waals forces, the machine will in some cycles extract *much* more heat than $Q = kT \ln 2 = (26)$ from the bath, and convert it into usable work. In other cycles, the machine will not convert heat into useful work, but convert work into heat, and dissipate that heat into the bath. The essential thing is, that the effect of the fluctuating van der Waals forces will in any case be larger by several orders of magnitude than the tiny force due to the single rare-gas atom.

Furthermore the appearance of van der Waals forces between

the pistons and the partition reminds us, that the van der Waals force between the 1 rare-gas atom, which drives Szilard's machine, and the vessel walls must not be ignored. Due to the induced electric dipole moment, the atom will quite often be adsorbed to the surfaces of vessel, partition, and pistons, and not at all behave like an ideal gas.

We made a further oversimplification, when we assumed that the gas pressure of the vessel walls and the pistons is zero. Of course the gas pressure of the solid surfaces of many metals is so small, that it is not measurable, not even at 1000 K. But in a consistent description of a machine, whose functionality rests on the tiny pressure of a 1-atom-gas, we certainly must not ignore the metal atoms, which sometimes will detach from the metal surfaces, drift through the vacuum, and demolish the functionality of Szilard's machine.

If we want to build a consistent model with the vessel, the partition, and the pistons as rigid classical bodies, and van der Waals forces being negligible, then we must not fill only one atom into the machine, but a sufficiently large amount of atoms as sketched in fig. 4.

Considering the diagram fig. 3, less than something like $N \approx 10^6$

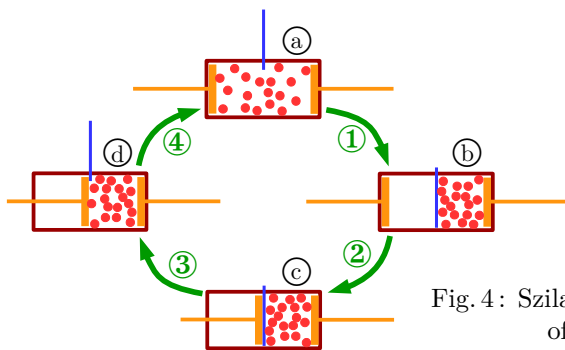


Fig. 4: Szilard's perpetual motion machine of second kind, corrected

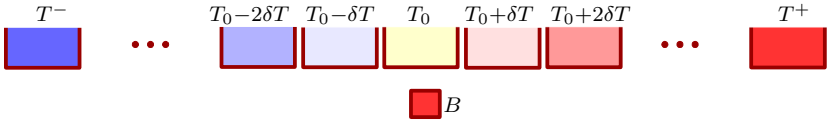
atoms would hardly be sufficient to make the van der Waals forces negligible. But then it becomes quite unlikely that in process step ①, when the partition is shifted-in, the numbers of atoms in the left and right cell significantly differ. Only in 2 of $2^{10^6} \approx 10^{18}$ cycles will all 10^6 atoms be trapped by chance in one cell, and the other cell will be empty. If the machine runs at 1 cycle per second, then the mean waiting time for that lucky event is about $18 \cdot 10^9$ years.

Of course we don't need that perfect event. A cycle with — say — $6 \cdot 10^5$ atoms in one cell and $4 \cdot 10^5$ atoms in the other may still allow for extraction of a little bit of work. The essential thing is that with this type of considerations, the second law turns from a matter of principle to a matter of probability. The second law in this understanding doesn't say “an isothermal cyclic machine will not extract heat from a bath and convert it into usable work in any cycle”. Instead it says “an isothermal cyclic machine will not extract a measurable amount of heat from a bath and convert it into a measurable amount of usable work in almost all cycles”, and statistical thermodynamics give a precise quantitative value for the rare exceptions.

With the probabilistic interpretation of entropy, there exists absolutely no problem with Szilard's machine fig. 4. It's merely unlikely, but not impossible, that the number of atoms trapped by chance in process step ① in one cell differs significantly from the number of atoms trapped in the other cell.

Once we accept that the second law may be violated in rare events, Szilard's perpetuum mobile turns into an argument *against* his theorem (12a), due to the absurd consequences at which we arrived when we assumed the theorem to be true.

3.3. The thermal memory



The third device, by which Szilard tried to demonstrate a relation between information and thermodynamic entropy, is an instrument which can measure the property Z of an object. This property can assume only two values: Z^+ or Z^- . If the instrument observes Z^+ , then a body B is first thermally contacted to a heat bath of temperature T_0 , then to a bath of temperature $T_0 + \delta T$, then to a bath of temperature $T_0 + 2\delta T$, and so on, and eventually to a bath of temperature T^+ . When the body B has assumed the temperature T^+ , it is thermally isolated from the baths, and the measurement is completed.

If the instrument observes Z^- , then the body B is first thermally contacted to the heat bath of temperature T_0 , then in infinitesimal steps cooled down to T^- , and then thermally isolated from the baths.

In either case, the result (Z^+ or Z^-) of the measurement is after the measurement documented by the temperature (T^+ or T^-) of the body B . Szilard emphasized that the heating of the body from T_0 to T^+ , and the cooling of the body from T_0 to T^- , (i. e. the adjustment of the thermal memory to the new value to be stored), is done reversibly (i. e. with no change of entropy) due to the infinitesimal small steps δT , while the heating or cooling of the body from its previous temperature to T_0 , (i. e. the deletion of the previously stored value from the thermal memory), is done irreversibly (i. e. with entropy increase of the environment). In contrast, in the example of the perpetuum mobile of second kind, Szilard had assumed that the entropy increase would happen in

course of the measurement. Only half a century later, Bennett [10] proposed an explanation of Szilard’s perpetuum mobile of second kind, in which again the irreversible step — correlated with entropy increase of the environment — does *not* happen in process step ② (i. e. in course of measurement of the atom’s position), but only later when the user of the machine clears from his memory the result of the previous measurement. Thereby the absurd consequences of the entropy increase of the environment happening in course of the measurement, which have been mentioned in the previous section, can be avoided. Bennett’s point of view will be described in the [appendix](#).

Now Szilard entertained some rather intricate considerations, to demonstrate that this storage of measurement results increases the entropy of the baths by $\Delta S_{\text{bath}} = k \ln 2$ per 1-bit result, as postulated in his theorem (12a). Those considerations are — in my opinion — not worth the effort, because that setup of thermal baths is anyway quite arbitrary. The mechanism could for example be constructed such, that the body B never touches any heat bath, i. e. that merely the position of B codes for the result, with “ B nearby (but never touching) the bath T^+ ” coding for the measured result Z^+ , and “ B nearby the bath T^- ” coding for the measured result Z^- . Then the memory would still serve it’s purpose (storing the result until the next measurement is started), but the entropy change would be zero. Thereby this machine becomes an argument *against* Szilard’s theorem of an unavoidable entropy change upon measurements.

It’s probably fair to say that none of the three machines, which Szilard presented, can be acknowledged a sufficient argument for his theorem (12a). In case of the self-measuring gas, he arbitrarily confused information entropy and thermodynamic entropy. That logical error is called “begging the question”: Szilard used in

the derivation the alleged equivalence of information entropy and thermodynamic entropy, and thereby arrived at the result that information entropy and thermodynamic entropy are equivalent. In case of the perpetuum mobile, the assumption of thermodynamic entropy reduction in course of the measurement resulted in absurd consequences. And the thermal memory could easily be modified to an example against Szilard's theorem (12a).

The confusion is not surprising, as Szilard came up with his ideas 20 years before Shannon worked out a clear definition of information entropy. Szilard had the vague idea, that there was something important to be found out with regard to a relation between information and entropy. But he was not yet able to pinpoint the issue with sufficient precision.

4. Landauer's principle

If the input of a logical operation can be uniquely concluded from the output, then that operation is called "logically reversible". The hallmark of logically reversible operations is a unique one-to-one map between the input configurations and the output configurations. Logical operations, which channel several different input configurations into one output configuration are logically irreversible. Logically irreversible operations reduce the information entropy, while logically reversible operations don't.¹⁰

Here are two most simple examples:

INVERT		RESTORE TO ONE	
input	output	input	output
1	0	1	1
0	1	0	1

¹⁰ A device which channels one input configuration into several different output configurations, and thereby increases information entropy, would not be a functional logical device but something like a random number generator.

The INVERT operation is logically reversible. In contrast, from the output 1 of the RESTORE TO ONE operation, we can not conclude whether the input is 1 or 0.

Let's compute the changes $\Delta H = H_{\text{output}} - H_{\text{input}}$ of information entropy generated by these two operations. If we assume for the input data the probabilities $P(0) = P(1) = 1/2$, then we get

$$\begin{aligned} H(\text{INVERT,input}) &\stackrel{(11)}{=} -2 \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ bit} \\ H(\text{INVERT,output}) &\stackrel{(11)}{=} -2 \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ bit} \\ \implies \Delta H(\text{INVERT}) &= 0 \end{aligned} \tag{31a}$$

$$\begin{aligned} H(\text{RESTORE TO ONE,input}) &\stackrel{(11)}{=} -2 \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ bit} \\ H(\text{RESTORE TO ONE,output}) &\stackrel{(11)}{=} \frac{1}{1} \log_2 \frac{1}{1} = 0 \\ \implies \Delta H(\text{RESTORE TO ONE}) &= -1 \text{ bit} . \end{aligned} \tag{31b}$$

Rolf Landauer (1927–1999) published in 1961 an influential article [13], in which he asserted what became well-known under the name

Landauer's principle: If a logical operation reduces the information entropy of the processed data by

$$\Delta H = -1 \text{ bit} , \tag{32a}$$

then the thermodynamic entropy of the hardware, which implements that logical operation, is reduced by minimum

$$\Delta S = -k \ln 2 .$$

According to the second law of thermodynamics, the entropy reduction of the hardware must be compensated by an entropy increase of the environment of minimum same size:

$$\Delta S_{\text{bath}} \geq k \ln 2 \quad (32b)$$

Thus at temperature T minimum the heat²

$$-Q \geq kT \ln 2 \quad (32c)$$

must be dissipated, and consequently according to the first law minimum the work

$$W \geq kT \ln 2 \quad (32d)$$

must be done, to accomplish any logical operation which reduces the information entropy of the processed data by 1 bit.

The similarity of Landauer's principle and Szilard's theorem is striking. A minimum entropy reduction $\Delta S = -k \ln 2$ of the hardware is postulated

- * “per bit of measurement result for any measurement and storage of the result” in case of Szilard's theorem (12a), respectively
- * “per bit of information entropy reduction of the processed data” in case of Landauer's principle (32a).

As a measurement clearly is correlated with a reduction of information entropy (we get the information of the — previously unknown — actual value of a physical quantity), the theorems of Szilard and Landauer are closely related. An important difference is that Landauer had the precise notion “information entropy” (defined by Shannon in 1948) available, while Szilard relied on the somewhat vague notion “measurement”. Obviously Szilard's prior work was unknown to Landauer, as he did not cite it in his 1961 publication.

With regard to logically reversible operations, Landauer asserted in later publications, that

“**reversible classical computation** [...] can be accomplished with as little energy dissipation, per step, as desired.” [14] (33)

Thus — in his opinion — the entropy increase $\Delta S_{\text{bath}} \geq k \ln 2$ is strictly coupled to the information entropy reduction of 1 bit; as reversible logical operation don't reduce the information entropy, they don't induce an inevitable increase of thermodynamic entropy of the environment.

Lossless computation has been advocated in particular by Fredkin and Toffoli [15] and by Bennett [9]. Objections have been raised e. g. by Porod et. al. [16]. The question is controversial still by today.¹¹

Like Szilard, Landauer didn't state his theorem explicitly. Instead (32a) is a condensed form of arguments presented in section "4. Logical irreversibility and entropy generation" of his article [13]. In this section Landauer considers the example of the RESTORE TO ONE operation, and argues:

“Note that our argument here does not necessarily depend upon connections, frequently made in other writings, between entropy and information. We simply think of each bit as being located in a physical system, with perhaps a great many degrees of freedom, in addition to the relevant one. However, for each possible physical state which will be interpreted as a ZERO, there is a very similar possible physical state in which the physical system represents a ONE. [...]” (34)

Consider a statistical ensemble of bits in thermal equilibrium¹². If these are [in a RESTORE TO ONE operation]

¹¹ A “commercial” argument: High-power processors are a market of about 10^{11} US\$ per year, and the power dissipation of the devices is a predominant problem, severely impeding the progress of computer technology. Why, then, did no player in this market make — despite the huge 10^{11} US\$ incentive — the slightest progress towards loss-less computation, even though the idea for loss-less computation is available since more than half a century?

¹² From the context it is obvious, that Landauer is speaking of an ensemble of bits, in which ONE and ZERO are showing up with approximately same frequency. He is definitively *not* speaking of a system in thermal equilibrium,

all reset to ONE, the number of states covered in the ensemble has been cut in half. The entropy therefore has been reduced by $k \log_e 2 = 0.6931 k$ per bit. The entropy of a closed system, e. g., a computer with its own batteries, cannot decrease; hence this entropy must appear elsewhere as a heating effect, supplying $0.6931 kT$ per restored bit to the surroundings. This is, of course, a minimum heating effect, and our method of reasoning gives no guarantee that this minimum is in fact achievable.”

Landauer starts his explanation with the words “that our argument here does not necessarily depend upon connections, frequently made in other writings, between entropy and information.” But exactly that is, what he is actually doing: He implicitly assumes that information entropy and thermodynamic entropy are equivalent. While Szilard erratically sometimes made this assumption, but didn't make it at other places, Landauer consistently bases all his 1961 article (and all his later writings) on this basic assumption.

In his argument (34), Landauer clearly skipped the distinction between information entropy and thermodynamic entropy, which Shannon had not yet denied. Landauer's point of view since then has been adopted by many physicists, though certainly not by all. It has been worked out in particular by Bennett [9]. Landauer's arguments are not easy to understand. Fortunately there exists a particular clear explanation of Landauer's reasoning, published by M. P. Frank [17]. In the sequel, I will go along Frank's account of the issue.

Frank first of all defines information entropy slightly different from Shannon's original definition. Shannon [7] defined the in-

in which — depending on temperature — some certain percentage of the bits is set *by chance* to ONE and the other bits are set to ZERO. A computer can only work correctly well apart from thermal equilibrium, i. e. as long as it's bits are *not* in thermal equilibrium.

formation entropy of a variable C , which assumes the n different values c_j with probabilities $P(c_j)$, due to

$$H(C) \stackrel{(11)}{=} -K \sum_{j=1}^n P(c_j) \log_2 P(c_j) \quad (35)$$

$$\text{with } K = 1 \quad \text{and} \quad \sum_{j=1}^n P(c_j) = 1 .$$

Shannon had already emphasized that his choice 1 for the constant K is completely arbitrary. To give information entropy the same dimension (energy/temperature) as thermodynamic entropy, Frank decided for $K = k/(\log_2 e)$, resulting into

$$H(C) = -\frac{k}{\log_2 e} \sum_{j=1}^n P(c_j) \log_2 P(c_j) = -k \sum_{j=1}^n P(c_j) \ln P(c_j)$$

$$\text{with } \sum_{j=1}^n P(c_j) = 1 , \quad (36a)$$

thereby making it formally almost identical to Gibbs' definition

$$S(c_j) \stackrel{(8)}{=} -k \sum_i p(\phi_i^{(j)}) \ln p(\phi_i^{(j)}) \quad \text{with} \quad \sum_i p(\phi_i^{(j)}) = 1 \quad (36b)$$

for the thermodynamic entropy of an observable macrostate c_j which may be realized with probabilities $p(\phi_i^{(j)})$ by any of the not observable microstates $\phi_i^{(j)}$.

Consider a computer which can assume n computational macrostates c_j , which again may be realized each by any of a finite set of microstates $\phi_i^{(j)}$ of the hardware atoms in phase space:

$$\underbrace{\phi_1^{(1)}, \dots, \phi_m^{(1)}}_{c_1}, \underbrace{\phi_1^{(2)}, \dots, \phi_r^{(2)}}_{c_2}, \dots, \underbrace{\phi_1^{(n)}, \dots, \phi_s^{(n)}}_{c_n}$$

Thus the c_j define a discrete partition of the set of the microstates ϕ_i . Frank concludes [17, eq. (17)] that the following relation must hold between the probabilities $P(c_j)$ of the macrostates c_j and the probabilities $\tilde{p}(\phi_i^{(j)})$ of the microstates $\phi_i^{(j)}$:

$$P(c_j) = \sum_i \tilde{p}(\phi_i^{(j)}) \quad (37)$$

Note that Frank's $\tilde{p}(\phi_i^{(j)})$ are different from Gibbs' $p(\phi_i^{(j)})$:

$$\tilde{p}(\phi_i^{(j)}) \stackrel{(36b),(37)}{=} p(\phi_i^{(j)}) \cdot P(c_j) \quad (38)$$

Now the following definition of *total entropy* seems reasonable:

$$\begin{aligned} S(\Phi) &= -k \sum_i \sum_j \tilde{p}(\phi_i^{(j)}) \ln \tilde{p}(\phi_i^{(j)}) = \\ &\stackrel{(38)}{=} -k \sum_i \sum_j p(\phi_i^{(j)}) P(c_j) \left(\ln P(c_j) + \ln p(\phi_i^{(j)}) \right) = \\ &= -k \underbrace{\sum_j \sum_i \overbrace{p(\phi_i^{(j)})}^1 P(c_j)}_{H(C)=(36a)} \ln P(c_j) + \\ &\quad + \underbrace{\sum_j P(c_j) \left(-k \sum_i \overbrace{p(\phi_i^{(j)}) \ln p(\phi_i^{(j)})}^{S(c_j)=(8)} \right)}_{S(\Phi|C)} \end{aligned} \quad (39)$$

Frank comments [17, below eq. (18)]: “In other words, the (total) physical entropy $S(\Phi)$ is *exactly equal* to the information entropy $H(C)$ of the computational state, plus the conditional entropy $S(\Phi|C)$ of the physical state, conditioned on the computational

state — this just means [...] the entropy that we would expect the physical state Φ to still have, if we were to learn the exact value of the computational state C .”

Frank names (39) the *Fundamental Theorem of the Thermodynamics of Computation*. I think it should not be called a theorem, but a definition, because this is exactly what it is: The definition (implicitly due to Landauer 1961) of a new notion of entropy, in which thermodynamic entropy as defined in (8) by Boltzmann and Gibbs, and Shannon’s information entropy (36a), are amalgamated to just one overall type of entropy, which may appropriately be called total entropy.

(39) becomes part of a theorem, however, due to the claim (which Landauer made) that the total entropy $S_{(39)}$ as defined in (39) exactly matches Clausius’ thermodynamic entropy $S_{(3)}$ as defined in (3), and that consequently a reduction $\Delta S_{(39)} < 0$ implies — due to the second law — the emission of the heat²

$$-\Delta Q \stackrel{(6)}{=} -T\Delta S_{(39)} = +T\Delta S_{\text{bath}} > 0 \quad (40a)$$

from the system to the environment. The problem is, that Boltzmann and Gibbs raised the same claim for their entropy $S_{(8)}$ as defined in (8), i. e. that a reduction $\Delta S_{(8)} < 0$ implies the emission of the heat²

$$-\Delta Q \stackrel{(6)}{=} -T\Delta S_{(8)} = +T\Delta S_{\text{bath}} > 0 \quad (40b)$$

from the system to the environment. As in general $\Delta S_{(39)} \neq \Delta S_{(8)}$, both claims (40) can impossibly be right (they could of course both be wrong).

The only way to disprove (40a) and/or (40b) is by experiment. To understand why such a decisive experiment is extremely difficult, let’s evaluate eq. (39) for the example of a RESTORE TO ONE

operation. As Landauer assumed that ZERO and ONE bits come in with same frequency, the total entropy $S_{(39)}$ of the input is

$$\begin{aligned}
 S_{(39), \text{in}} &\stackrel{(39)}{=} \underbrace{\left(\frac{1}{2} + \frac{1}{2} \right) k \ln 2}_{H(C)} + \\
 &+ \frac{1}{2} \underbrace{\left(-k \sum_i p(\phi_i^{(\text{ZERO in})}) \ln p(\phi_i^{(\text{ZERO in})}) \right)}_{S_{(8)}(\text{ZERO in})} + \\
 &+ \frac{1}{2} \underbrace{\left(-k \sum_r p(\phi_r^{(\text{ONE in})}) \ln p(\phi_r^{(\text{ONE in})}) \right)}_{S_{(8)}(\text{ONE in})} .
 \end{aligned}$$

The total entropy $S_{(39)}$ of the output is

$$\begin{aligned}
 S_{(39), \text{out}} &\stackrel{(39)}{=} \underbrace{-k \left(\overbrace{0 \ln 0}^0 + \overbrace{\ln 1}^0 \right)}_{H(C)} - \\
 &\underbrace{-k \sum_s p(\phi_s^{(\text{ONE out})}) \ln p(\phi_s^{(\text{ONE out})})}_{S_{(8)}(\text{ONE out})} .
 \end{aligned}$$

Thus the change of total entropy $S_{(39)}$ upon the RESTORE TO ONE operation is

$$\begin{aligned}
 \Delta S_{(39)}(\text{RESTORE TO ONE}) &= \overbrace{-k \ln 2}^{\Delta H(\text{RESTORE TO ONE})} + \\
 &+ \underbrace{S_{(8)}(\text{ONE out}) - \frac{1}{2} S_{(8)}(\text{ONE in}) - \frac{1}{2} S_{(8)}(\text{ZERO in})}_{\Delta \bar{S}_{(8)}(\text{RESTORE TO ONE})} . \quad (41a)
 \end{aligned}$$

Note that

$$\Delta S(\Phi|C) = \overline{\Delta S}_{(8)} ,$$

i. e. the change $\Delta S(\Phi|C)$ of the conditional entropy of the physical state is identical to the mean change $\overline{\Delta S}_{(8)}$ of thermodynamic entropy as defined by Boltzmann and Gibbs, see (39). Obviously Landauer assumed that $\overline{\Delta S}_{(8)}$ would be zero upon both logically reversible and logically irreversible logical operations, because only then his two claims (32a) and (33) result from (41a) and from

$$\begin{aligned} \Delta S_{(39)}(\text{INVERT}) &= \overbrace{0}^{\Delta H(\text{INVERT})} + \\ &+ \frac{1}{2} \left(S_{(8)}(\text{ONE out}) + S_{(8)}(\text{ZERO out}) - \right. \\ &\quad \left. - S_{(8)}(\text{ZERO in}) - S_{(8)}(\text{ONE in}) \right) . \end{aligned} \quad (41b)$$

$$\underbrace{\hspace{15em}}_{\overline{\Delta S}_{(8)}(\text{INVERT})}$$

$\overline{\Delta S}_{(8)} = 0$ is certainly *not* to be expected, however, with the usual transistor implementations of logical operations in computers. The numbers of microstates, which are identified as a ZERO or ONE may be, say, something like $(5 \pm 1) \cdot 10^{20}$. Then we have

$$\begin{aligned} \overline{\Delta S}_{(8)}(\text{RESTORE TO ONE}) &\approx \overline{\Delta S}_{(8)}(\text{INVERT}) \approx \\ &\approx (-1 \dots + 1)k \ln 10^{20} \approx (-1 \dots + 1) \cdot 67k \ln 2 \approx \\ &\approx (+1 \dots - 1) \cdot 67 \cdot \Delta H(\text{RESTORE TO ONE}) . \end{aligned}$$

According to this very rough estimation, the range of uncontrolled variations of $\overline{\Delta S}_{(8)}$ would be more than 130× as large as the change ΔH of information entropy. But $\overline{\Delta S}_{(8)}$ must be reliably controlled down to a possible error $\ll k$ for a significant check of (40), because $|\Delta H|$ is by order of magnitude typically $\mathcal{O}(k)$ for a logical operation.

Thus a special hardware implementation of the logical operations is required, to make the small contribution to the heat dissipation — if it should exist — of ΔH visible.

Besides the problem to eliminate the contribution of $\overline{\Delta S}_{(8)}$, it's anyway a very difficult task to approach the limit

$$\Delta S = -k \ln 2 \quad (42a)$$

in an experimental realization of logical 1-bit-processes, and measure (respectively conclude from other experimental parameters) the related tiny work

$$W = kT \ln 2 \approx 3 \cdot 10^{-21} \text{J} \quad \text{at room temperature ,} \quad (42b)$$

which must be done to drive the process. In the sequel several experiments are discussed, which have been performed to check Landauer's principle.

4.1. Experimental evaluations

Landauer's principle (32) states that logically irreversible logical operations imply a power dissipation by the hardware of minimum $kT \ln 2$ per bit of information entropy reduction. And in (33) Landauer asserts that no minimum power dissipation exists for logically reversible logical operations. Several experiments have been performed, which targeted to check these assertions.

4.1.1. A bead trapped in the laser focus

Berut et.al. [18] were the first to report an experiment of that type. The title “Experimental verification of Landauer's principle linking information and thermodynamics” of their publication, however, is misleading. Their experiment did not demonstrate whatever link between information and thermodynamics, and in

particular it did not verify Landauer’s principle (32). Instead their experiment demonstrated that the work, which had to be done to drive the thermodynamically irreversible process, which they analyzed, approached — with decreasing switching speed — from above the limit $kT \ln 2$.

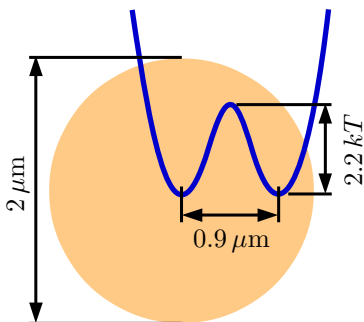


Fig. 5: The silica bead in the double-well potential

In their experiment, a silica bead of $2 \mu\text{m}$ diameter, which was suspended in water, was trapped in the focus of a laser beam. The laser beam was focused onto two nearby points with a switching rate of 10 kHz , such that a double-well potential as sketched in fig. 5 was created, with distance $0.9 \mu\text{m}$ between the two minima, and a barrier height of about $2.2 kT$ between the two wells.

At the begin of each experimental run, the bead was in one well, the other well was empty. Even though the bead diameter was much larger than the distance of the minima of the two wells, the position of the bead could still be clearly assigned to one of the wells, because it was observed at 502 Hz with a fast camera, which allowed (after picture treatment) for a resolution of better than 10 nm .

Then the water cell was moved for a time τ with speed v relative to the laser focus. The speed was during the time interval τ linearly increased from $v = 0$ to $v = v_{\text{max}}$. Due to the viscous drag of the

water, the bead felt the force

$$F = \gamma \cdot v \quad (43a)$$

$$\gamma = 1.89 \cdot 10^{-10} \text{ Ns/m} = \text{coefficient of friction} ,$$

which was driving the bead towards the empty well. This process can of course be interpreted as a realization of the logically irreversible RESTORE TO ONE operation, with “bead in the left well” coding for ZERO, and “bead in the right well” coding for ONE. Note, however, that this process can as well be interpreted as a realization of the logically reversible INVERT operation.

The time τ was varied in the range from 5 s to 40 s, and the work

$$W(\tau) = \int_0^{\tau} dt F(t) v(t) \quad (43b)$$

done onto the bead was computed. It turned out that $W(\tau)$ is an exponentially decreasing function, which does not converge against zero, but against

$$\lim_{\tau \rightarrow \infty} W(\tau) = kT \ln 2 . \quad (43c)$$

4.1.2. Clearing an RC-memory cell

Orlov et. al. [19] tried to check Landauer's principle by means of a surprisingly simple setup: The memory sketched in fig. 6 is by definition set to ONE with $U_C \approx +50 \mu\text{V}$, it is set to ZERO

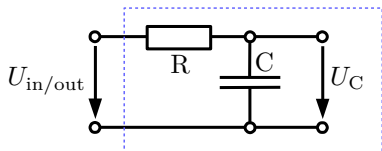


Fig. 6: The RC-memory cell

with $U_C \approx -50 \mu\text{V}$, and it is set to NEUTRAL with $U_C \approx 0\text{V}$. With $C = 100\text{pF}$, the bit-energy is $C \cdot (50 \mu\text{V})^2/2 \approx 30 kT$ at room temperature. Thus the memory is stable against thermal fluctuations, which are $\mathcal{O}(kT)$.

Two methods to clear the memory were evaluated:

In the method “clear without memory”, $U_{\text{in}} = 0\text{V}$ was applied to the input clamps. Thereby the bit energy of $30 kT$ was dissipated in the resistor R .

In the method “clear with memory”, first the actually stored value (either ONE, coded by $U_C \approx +50 \mu\text{V}$, or ZERO, coded by $U_C \approx -50 \mu\text{V}$) was measured, and the measurement result stored in the brain of the experimenter, i. e. a copy of the memory content was prepared in the experimenter’s brain. Then an external, continuously variable voltage source was connected to the input clamps, with the value of the voltage source adjusted exactly to the previously measured memory voltage. Then the external voltage was slowly ramped down or up to 0V . With this method, the power dissipation in the resistor R can be made arbitrarily small, if the external voltage is tuned sufficiently slowly. With a total switching time of $670 \mu\text{s}$ from $\pm 50 \mu\text{V}$ to 0V , the measured power dissipation over R was as low as $0.01 kT$, i. e. less than 1‰ of the $30 kT$ bit energy.

Orlov et. al. concluded: Due to the previous measurement of the actual memory value, and storage of the measurement result in the experimenter’s brain, “clear with memory” is a reversible operation. Indeed, the memory could be easily reset to the previous voltage by tuning the external power supply from 0V to the previously measured voltage. In contrast, “clear without memory” is a irreversible operation: As nobody knows what value had been stored in the memory before the erasure, that value is lost forever.

In the opinion of Orlov et. al. these results confirm Landauer’s assertions (32a) and (33) regarding the power dissipation of ir-

reversible and reversible logical operations. Thus they obviously believed that

$$\overline{\Delta S}_{(8)}(\text{CLEAR}) \stackrel{(41)}{=} S_{(8)}(U_C = 0) - \frac{1}{2} S_{(8)}(U_C = +50 \mu\text{V}) - \\ - \frac{1}{2} S_{(8)}(U_C = -50 \mu\text{V}) \stackrel{?}{\ll} k,$$

would be much smaller than the Boltzmann constant k . But such precise balance of the three $S_{(8)}$ -values seems unbelievable, given that the number of different microstates, which all are interpreted as one certain memory value, probably is something like $\approx 10^{20}$, given that the capacitor is a macroscopic device.

Furthermore a trivial — and fatal! — error has slipped into the reasoning of these authors: How is it possible that the variable voltage source supplies a stable voltage with no significant thermal fluctuations? This is *only possible* because the source is dissipating in its internal regulation circuitry *much more* energy than kT .

The voltage source is an integral, indispensable part of the experimental setup, which must not be ignored. Orlov et. al. could as well perform the “clear without memory” operation by shortening the input clamps with an external resistor R_{ext} . Then only $R/(R + R_{\text{ext}})$ of the bit energy would be dissipated in R , while $R_{\text{ext}}/(R + R_{\text{ext}})$ of the bit energy would be dissipated in the external resistor. If it were correct to ignore the external energy dissipation, then the energy dissipation in the internal resistor R could be made arbitrarily small by choosing R_{ext} arbitrarily large. But that's of course plain nonsense. The external energy dissipation, be it in the external voltage source or be it in the external resistor R_{ext} , must clearly *not* be ignored.

4.1.3. A particle in the electrokinetic trap

Jun et. al. [20] trapped a 200 nm fluorescent particle not in the focus of a laser, but in the electric potential created by a “electrokinetic feedback trap”. This is a trap, in which a varying voltage is applied to electrodes. Depending on the observed position of the bead, the voltage was chosen such that the bead felt a double-well potential with variable height of the intermediate barrier. The update time of the feedback loop was 10 ms.

Two types of processes were run: In process ①, the bead was shifted (due to frictional forces of the water) from that well, where it was first observed, to the other well, thereby simulating the logically irreversible RESTORE TO ONE operation, or the logically reversible INVERT operation. In process ②, the intermediate barrier was lowered, and after some time raised again, without exerting frictional forces due to the water. In this process, the bead ended up in either well with probability $1/2$. This operation may be called RANDOMIZE. It is a defining requirement of any logical operation, that the output data are uniquely determined by the input data. The RANDOMIZE operation does not meet this requirement, hence it is no logical operation at all.

Again it turned out that the (computed) work, which was done on the particle, was lower, if the process was run slower. The extrapolated computed work done on the particle was

$$\lim_{\tau \rightarrow \infty} W(\tau) = \begin{cases} kT \ln 2 & \text{with process ①} \\ 0 & \text{with process ②.} \end{cases} \quad (44)$$

Shifting a particle through a viscous fluid is by construction a thermodynamically irreversible process. Jun et. al. demonstrated, that a logical operation, if realized by a thermodynamically irreversible process, needs minimum the work $kT \ln 2$ per processed bit of information.

Actually Jun et. al. interpreted their results as support for Landauer's principle (32), because no energy was dissipated in process ②. But this interpretation is clearly mistaken:

In process ①, work was done to overcome the potential barrier and shift the bead into the target well. In process ②, the experimenters simply leaned back, and let the bead drift into whatever well it wanted to drift. Thus quite trivially work had to be done in process ①, and no work needed to be done in process ②.

Process ②, which gave the $\lim_{\tau \rightarrow \infty} W(\tau) = 0$ result, is realizing the RANDOMIZE operation, which is no logical operation. Process ①, which gave the $\lim_{\tau \rightarrow \infty} W(\tau) = kT \ln 2$ result, is the (thermodynamically irreversible) realization of both the logically reversible INVERT operation and the logically irreversible RESTORE TO ONE operation. Hence Jun et. al. did *not* demonstrate that a logical operation (be it logically reversible or irreversible) can be performed with no energy dissipation.

4.1.4. Switching a nanomagnetic memory

A further experiment, which deserves mention, has in 2016 been published by Hong. et. al. [21]. These experimenters measured the (magnetic) work needed to switch 1-domain nanomagnetic memories of lateral dimensions slightly below 100 nm. The nanomagnets had an “easy” axis and a “hard axis”. More energy is needed to align the magnetic moment to the hard axis than to the easy axis. Hence in thermal equilibrium the magnetic moment is aligned either parallel (this codes for ZERO) or antiparallel (this codes for ONE) to the easy axis.

To switch the memory from ZERO to ONE or from ONE to ZERO, work must be done due to externally applied magnetic fields. Hong. et. al. measured the work needed to switch the memory, and found values which come close to $W \approx kT \ln 2$, but scattered

by 50 % to 100 % around this value.

Hong. et. al. interpreted this result as a confirmation of the Landauer limit $W \geq kT \ln 2$. As their process, different from the bead shifted against water friction, is not obviously dissipative, a different speculation is tempting: Of course we must do work to turn the magnetic moment from the easy axis into the hard axis. But why don't we get this work 100 % back, when the moment is relaxed from the hard axis to it's new orientation along the easy axis? Indeed the many results with W well below $kT \ln 2$ might indicate (??) that this process might ideally be lossless (i. e. thermodynamically reversible), and that the results $W > 0$ might be caused by experimental imperfections. If, on the other hand, the result $W \geq kT \ln 2$ should be confirmed for this system, then this power dissipation must be caused by thermal fluctuations, but not by ordinary friction.

Thus it is a pity that the results of just this experiment are relatively unclear, and do not really confirm or disprove the importance of thermal fluctuations.

4.2. The impact of thermodynamic fluctuations

None of the experiments reported until today could prove (40a) and/or (40b) wrong. Actually a successful experiment may be impossible, due to the impact of thermodynamic fluctuations.

To see how thermodynamic entropy is affected by fluctuations, look for example at this experiment [22, 23], sketched in fig. 7 on the next page: Particles of $6 \mu\text{m}$ diameter were held in the focus of a laser, and dragged with velocity v through an aqueous solvent. (Note the similarity of this experiment with the experiments described in sections 4.1.1 and 4.1.3!). If integrated over long times (several seconds), the bead lagged the center of the potential, as sketched in fig. 7(a). This means that the bead-laser system

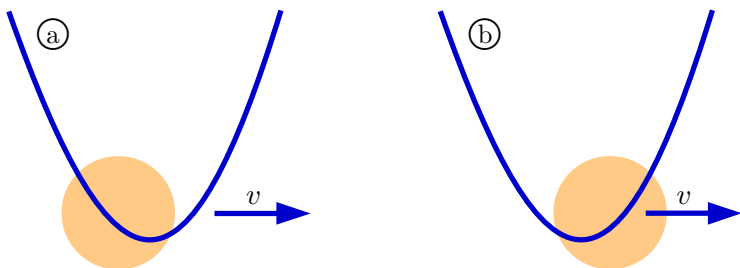


Fig. 7: Observing entropy fluctuations: The blue potential (created by a laser focus) is dragging a bead with velocity v through an aqueous solution.

dissipated heat to the solvent (i. e. the entropy of the solvent was increased), because the laser had to do work against the frictional force exerted by the solvent onto the bead. But when the system was observed with fine time resolution (e. g. 0.01 seconds), then in a significant part of the observation intervals the bead had to be slowed down by the laser focus, as sketched in fig. 7 (b). This means that sometimes for short time intervals heat was absorbed from the solvent (i. e. the entropy of the solvent was reduced), and converted into macroscopic kinetic energy of the bead.

Actually that experiment observed essentially the same effect, which Brown observed and published [24] — but was not yet able to understand — already in 1828, i. e. the erratic “Brownian motion” of tiny (just visible in the optical microscope) particles suspended in water. Only in 1905 Einstein [25] presented the correct explanation: Due to rare, but not impossible spontaneous fluctuations of thermodynamic entropy, a significant share of the water molecules nearby the particle, which normally move fully disordered (thermodynamic equilibrium), by chance moves into one common direction (spontaneous decrease of thermodynamic entropy), transferring macroscopic visible momentum to the particle. Thus in this phe-

nomenon *chance* is doing, what *intentionally* only Maxwell's demon is able to do: To cool a part of the water to a temperature which is lower than the temperature of the surrounding water (and lower than the temperature of the macroscopic particle, which is in thermodynamic equilibrium with the surrounding water), thereby transforming some heat of the water into macroscopic kinetic energy.

Observations like the fluctuating bead entropy [22, 23] or Brownian motion [24] disprove Clausius' assumption that the thermodynamic entropy of a system can *never* decrease, unless that decrease is compensated (or over-compensated) by an increase of thermodynamic entropy of the environment. But they are in full accord with Boltzmann's and Gibbs' understanding, that uncompensated small entropy fluctuations are happening all the time, while uncompensated observable huge fluctuations are utterly unlikely (but not impossible!) in the $> 10^{20}$ -particles-systems which these scientists considered. To make notions like small or huge quantitative, we might — somewhat arbitrary — say that spontaneous reductions of thermodynamic entropy of a closed system (which are violating the deterministic second law of Clausius' conception, but are in accord with the probabilistic second law of Boltzmann's and Gibbs' conception) by

- * $< k/2$ happen very frequently
- * $k/2 \dots 2k$ happen frequently
- * $2k \dots 5k$ happen sometimes
- * $5k \dots 20k$ happen rarely
- * $> 20k$ happen very rarely.

Consequently we will run into nonsense results, if we forget to take spontaneous fluctuations of thermodynamic entropy into account in evaluations of systems, in which entropy changes of order of magnitude $\mathcal{O}(k)$ are *not* negligible. For this reason, an experimental decision between the incompatible entropy concepts

(39), in which information entropy and thermodynamic entropy are amalgamated to a “total entropy” as implicitly defined by Landauer, and

(8)&(11), in which thermodynamic entropy as defined by Boltzmann and Gibbs, and information entropy as defined by Shannon, are kept separated,

may very well be impossible forever, because the difference between both concepts is $\mathcal{O}(k)$, hence masked by the unavoidable fluctuations of thermodynamic entropy.

4.3. Conclusions

One might argue — given the experimental situation described in the previous section — that a preference for one of the two concepts is merely a matter of taste. But I think that there are minimum four important arguments in favor of the (8)&(11) concept with two different types of entropy kept separated:

- * With the “total entropy” concept (39), the small changes ΔH of information entropy get in almost all (or all?) cases masked by the much larger changes $\Delta S(\Phi|C)$ of conditional total entropy (which are identical to the mean changes $\overline{\Delta S}_{(8)}$ of thermodynamic entropy as defined by Boltzmann and Gibbs). Whenever in successful evaluations of communication theory allegedly the concept (39) has been applied, actually the changes $\Delta S(\Phi|C)$ of conditional total entropy have been simply ignored, but not seriously been proved to be precisely zero or at least $\ll k$.
- * The separate concept of information entropy is free of spontaneous fluctuations. Hence changes of information entropy can with unlimited precision be traced down to $\Delta H = 0$ processes. This important advantage gets lost with the total entropy concept (39), which is unavoidably affected by the fluctuations

of thermodynamic entropy.

- * In the concept of a separate information entropy, there exists no counterpart for the second law of thermodynamics. There is no restriction for $\Delta H < 0$ processes.
- * The second law of thermodynamics is about the exchange of heat between the considered system and its environment. And in the Boltzmann/Gibbs conception, the thermodynamic entropy is proportional to the logarithm of the number of microstates in the phase space of the atomic constituents of the considered system, which all are compatible with the observed macrostate. The microstates in phase space determine — by assumption of the kinetic theory of heat — the heat content and temperature of the considered system.

The total entropy concept (39), on the other hand, assumes that ΔH has, in addition to $\overline{\Delta S}_{(8)}$, some impact on the heat exchange between system and environment. This assumption seems not to be motivated by a physical argument, to say the least, because only the phase space of the atomic constituents of the considered system, but not information space, is conceptually related to heat and temperature.

The thermodynamic entropy of system & environment increases *if* — and *because* — system & environment approach thermal equilibrium. The states which code in a computer for ZERO and ONE, in contrast, are *never* in thermal equilibrium with the environment, and they *do not approach* thermal equilibrium, as long as the computer is working correctly, see footnote 12. Thus it seems physically not well-founded to postulate whatever correlation between changes of these states and heat exchange with the environment.

Imagine that the computer is performing some certain operation at room temperature (25° Celsius). This operation results into the change

$$\Delta S_{(39)@25^{\circ}\text{C}} \stackrel{(41)}{=} \Delta H_{@25^{\circ}\text{C}} + \underbrace{\Delta S(\Phi|C)_{@25^{\circ}\text{C}}}_{\overline{\Delta S}_{(8)@25^{\circ}\text{C}}}$$

of total entropy. If we take the computer into a cooled room (15° Celsius) and perform the same operation, then

$$\Delta S_{(39)@15^{\circ}\text{C}} \stackrel{(41)}{=} \Delta H_{@15^{\circ}\text{C}} + \underbrace{\Delta S(\Phi|C)_{@15^{\circ}\text{C}}}_{\overline{\Delta S}_{(8)@15^{\circ}\text{C}}} .$$

If the computer works correctly at both temperatures, then clearly $\Delta H_{@15^{\circ}\text{C}} = \Delta H_{@25^{\circ}\text{C}}$, while obviously $\overline{\Delta S}_{(8)@15^{\circ}\text{C}} \neq \overline{\Delta S}_{(8)@25^{\circ}\text{C}}$, and consequently $\Delta S_{(39)@15^{\circ}\text{C}} \neq \Delta S_{(39)@25^{\circ}\text{C}}$. This shows again that ΔH is *not* related to temperature, hence *not* related to thermodynamic entropy, hence *not* subject to the second law.

While neither of the incompatible concepts (39) versus (8)&(11) is disproved by experiment (and therefore “wrong”), I think that for these reasons it is clearly preferable to keep the concept (8) of thermodynamic entropy and the concept (11) of information entropy separate, and *not* to mix them into the concept (39) of total entropy.

Appendix:

Bennett’s account of Szilard’s perpetual mobile

Bennett’s interpretation [9, sect. 5]¹³ of Szilard’s perpetual mobile of second kind is a beautiful example for an application of the “total entropy” concept (39). At the same time, it is (like Szilard’s own interpretation, discussed in section 3.2) a warning:

¹³ Bennett published his interpretation also in a more popular fashion — hence less clear — in [10].

It demonstrates that we run into purely theoretical speculations far-off from physical reality, if we forget to take the unavoidable (quantum and thermodynamic) fluctuations into account.

Bennett added to Szilard's machine a memory, see fig. 8. The

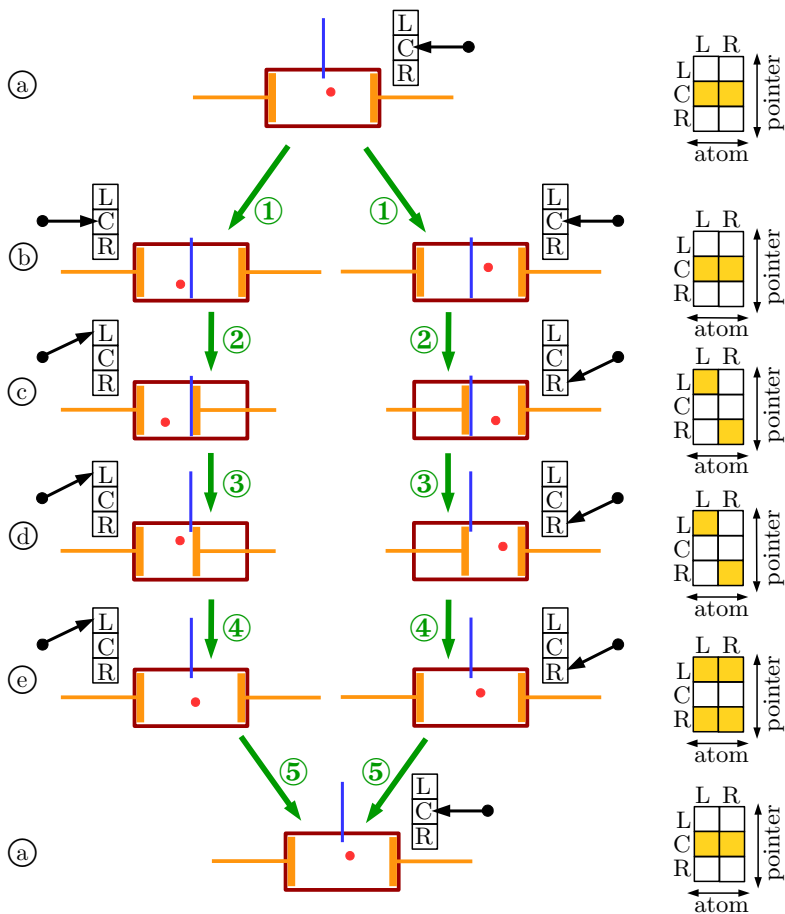


Fig. 8: Szilard's perpetual motion machine of second kind in Bennett's interpretation

pointer of the memory is set to L (R), if — due to process step ② — the atom is observed to be trapped in the left (right) half of the vessel. C is indicating the “clear” state of the memory. Note that Bennett added the memory only to make his point perfectly clear. He could as well skip the memory, and define “right (left) piston shifted in” as “L” (“R”) and “both pistons shifted out” as “C”.

In Boltzmann's and Gibbs' statistical model of thermodynamic entropy, the “phase space” of a system consisting of N hypothetical submicroscopic particles is the $6N$ -dimensional space of the $3N$ positions and $3N$ momenta of the N particles. Guided by (39), Bennett combines this phase space with the space of the informations stored in the memory to a combined space as sketched on the right side of fig. 8. Bennett names this combined space again “phase space”. To avoid confusion, I will stick to the conventional definition of “phase space”, and will use the name “combi-space” for the combination of phase space and information space.

In the combi-space diagrams of fig. 8, the left [right] column is marked by yellow color, if in the formula (39) of total entropy $p(\text{left}) > 0$ [$p(\text{right}) > 0$] because the left [right] half of the vessel is accessible to the atom. And the respective row(s) are marked, if with probability $P > 0$ the memory's pointer is indicating L and/or C and/or R, i. e. if in (39) $P(L) > 0$ and/or $P(C) > 0$ and/or $P(R) > 0$. All yellow marked squares, but not the white ones, add to the total entropy (39).

In state ① the atom is with probability $1/2$ in the left half of the vessel, and with probability $1/2$ in the right half of the vessel.

In state ② the atom is still with probability $1/2$ in the left half of the vessel, and with probability $1/2$ in the right half of the vessel. Thus there is in process step ① no change of yellow marked phase space volume, hence no change of total entropy.

Note that Bennett considers the two branches of fig. 8 in parallel, while Szilard assumed that in process step ① the atom is

trapped either in the left cavity EXOR in the right cavity, and consequently considered only one branch, see fig. 2 on page 22. The difference is due to the term $S(\Phi|C) = \sum_j P(c_j)S_{(8)}(c_j)$ in (39): Bennett considers $S(\Phi|C) = S_{(8)}(\text{left})/2 + S_{(8)}(\text{right})/2$, resulting into $\Delta S(\Phi|C) = \overline{\Delta S}_{(8)} = 0$ in process step ①. Szilard, in contrast, considers the actual entropy $S_{(8)}$, but not the mean entropy $\overline{S}_{(8)}$, resulting into $\Delta S_{(8)} = -k \ln 2$ in the same process step. In Bennett's account, the entropy reduction by $-k \ln 2$ is postponed to process step ⑤.

In process step ② the memory is set to L or R, depending on whether the right or the left piston gets shifted-in. As the probability for the result L or R is $1/2$ each, there is no change of combi-space volume, hence no change of total entropy. The yellow marks are merely shifted up or down in combi-space.


No change of accessible combi-space, hence no change of total entropy, happens in process step ③.

In step ④ the combi-space is doubled, increasing the machine's total entropy by $k \ln 2$. Therefore the machine can do in this step the useful work² $-kT \ln 2$.










State ⑥, however, is not yet identical to state ①. Instead now the memory must be reset to the "cleared" state C, to close the cycle, and this requires according to Landauer's principle the work² $+kT \ln 2$, thereby just compensating the useful work gained in the previous step. This compensation solves — according to Bennett [9, 10] — the problem of Szilard's perpetuum mobile of second kind: In total, per cycle no energy can be extracted, and the second law of thermodynamics is "saved".








Bennett's solution of Szilard's perpetuum mobile suffers, however, from the same problem as Szilard's own interpretation: Van der Waals forces must not be neglected but are dominating Szilard's machine (see fig. 3 on page 30). This makes Bennett's considerations anyway obsolete.









References

- [1] W. Thomson (Lord Kelvin): *On the Dynamical Theory of Heat*, Trans. Roy. Soc. Edinburgh, (March 1851),
 https://zapatopi.net/kelvin/papers/on_the_dynamical_theory_of_heat.html
- [2] R. Clausius: *Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie*, Ann. Phys. Chem. **201**, 353–400 (1865),
 <https://doi.org/10.1002/andp.18652010702>
 http://www.ngzh.ch/archiv/1865_10/10_1/10_3.pdf
- [3] Rudolf Clausius: *The Mechanical Theory of Heat — with its Applications to the Steam Engine and to Physical Properties of Bodies* (John van Voorst, London, 1867)
 <https://books.google.de/books?id=8LIEAAAAYAAJ>
- [4] L. Boltzmann: *Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht*. Wien. Ber. **76**, 373–435 (1877)
 auch in: L. Boltzmann: *Wissenschaftliche Abhandlungen Band II* (J. A. Barth, Leipzig, 1909)
 <https://phaidra.univie.ac.at/view/o:63651>
- [5] J. W. Gibbs: *On the Equilibrium of Heterogeneous Substances*, Connecticut Acad. Sci. (1875–77)
 I:  <https://archive.org/details/Onequilibriumhe00Gibb>
 II:  <https://archive.org/details/Onequilibriumhe00GibbA>
- [6] G. Gründler: *The three types of entropy*, APIN Circular se09113 (2018)
 <http://www.astrophys-neunhof.de/mtlg/se09113.pdf>

- [7] C. E. Shannon: *A Mathematical Theory of Communication*, Bell Syst. Tech. J. **27**, 379 – 423 + 623 – 656 (1948),
 <http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>
- [8] L. Szilard: *Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen*, Zeits. Phys. **53**, 840 – 856 (1929),
 <https://doi.org/10.1007/BF01341281> oder
 https://library.ucsd.edu/dc/object/bb1264347m/_1.pdf
English translation: *On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings*, Behav. Sci. **9**, 301 – 310 (1964),  <https://doi.org/10.1002/bs.3830090402>, alternative english translation:
 <http://www.sns.ias.edu/~tlusty/courses/InfoInBio/Papers/Szilard1929.pdf>
- [9] C. H. Bennett: *The thermodynamics of computation – a review*, Int. J. Theor. Phys. **21**, 905 – 940 (1982),
 <https://doi.org/10.1007/BF02084158> or  http://www.pitt.edu/~7Ejdnorton/lectures/Rotman_Summer_School_2013/thermo_computing_docs/Bennett_1982.pdf
- [10] C. H. Bennett: *Demons, Engines and the Second Law*, Sci. Am. **257**, 108 – 116 (1987),
 <https://ecee.colorado.edu/~ecen4555/SourceMaterial/DemonsEnginesAndSecondLaw87.pdf>
- [11] G. Gründler: *Zero-Point Energy and Casimir-Effect*, APIN Circular se08011 (2013)
 <http://www.astrophys-neunhof.de/mtlg/se08011.pdf>

- [12] F. London: *Zur Theorie und Systematik der Molekularkräfte*, Z. Phys. **63**, 245–279 (1930)
 <http://dx.doi.org/10.1007/BF01421741>
- [13] R. Landauer: *Irreversibility and Heat Generation in the Computing Process*, IBM J. R.&D. **5**, 183–191 (1961),
 <http://dx.doi.org/10.1147/rd.53.0183> or
 <http://worrydream.com/refs/Landauer - Irreversibility and Heat Generation in the Computing Process.pdf>
- [14] R. Landauer: *The physical nature of information*, Phys. Lett. A **217**, 188–193 (1996),  [https://doi.org/10.1016/0375-9601\(96\)00453-7](https://doi.org/10.1016/0375-9601(96)00453-7) or  http://cqi.inf.usi.ch/qic/64_Landauer_The_physical_nature_of_information.pdf
- [15] E. Fredkin, T. Toffoli: *Conservative logic*, Int. J. Theor. Phys. **21**, 219–253 (1982),
 <https://doi.org/10.1007/BF01857727>
- [16] W. Porod, R. O. Grondin, D. K. Ferry, G. Porod: *Dissipation in Computation*, Phys. Rev. Lett. **52**, 232–235 (1984),
 <https://doi.org/10.1103/PhysRevLett.52.232>
- [17] Michael P. Frank: *Physical Foundations of Landauer's Principle* In: J. Kari, I. Ulidowski (eds): *Reversible Computation. RC 2018*, Springer Lect. Notes Comp. Sci. **11106**
 https://doi.org/10.1007/978-3-319-99498-7_1
 <https://arxiv.org/abs/1901.10327>

- [18] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, E. Lutz : *Experimental verification of Landauer's principle linking information and thermodynamics*, Nature **483**, 187–189 (2012),
 <https://doi.org/10.1038/nature10872>  http://www.academia.edu/8720781/Experimental_verification_of_Landauer_s_principle_linking_information_and_thermo_dynamics
- [19] A. O. Orlov, C. S. Lent, C. C. Thorpe, G. P. Boechler, G. L. Snider : *Experimental Test of Landauer's Principle at the Sub – $k_B T$ Level*,
Jap. J. App. Phys. **51**, 06FE10 (5 pages) (2012),
 <https://doi.org/10.1143/jjap.51.06fe10>
 <https://www3.nd.edu/~lent/pdf/nd/OrlovExperimentalTestLandauerJJAP.pdf>
- [20] Y. Jun, M. Gavrilov, J. Bechhoefer : *High-precision test of Landauer's principle in a feedback trap*,
Phys. Rev. Lett. **113**, 190601 (2014),
 <https://doi.org/10.1103/PhysRevLett.113.190601>
 <https://arxiv.org/abs/1408.5089>
- [21] J. Hong, B. Lambson, S. Dhuey, J. Bokor : *Experimental test of Landauer's principle in single-bit operations on nanomagnetic memory bits*, Sci. Adv. **2**, e1501492 (2016),
 <https://doi.org/10.1126/sciadv.1501492>

- [22] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, D. J. Evans: *Experimental Demonstration of Violations of the Second Law of Thermodynamics for Small Systems and Short Time Scales*, Phys. Rev. Lett. **89**, 050601, 1–4 (2002),
 <https://doi.org/10.1103/PhysRevLett.89.050601> or
 <https://pdfs.semanticscholar.org/8864/bfc2771bed466520434f90d7ef0868f1fc20.pdf>
- [23] D. M. Carberry, J. C. Reid, G. M. Wang, E. M. Sevick, D. J. Searles, D. J. Evans: *Fluctuations and Irreversibility: An Experimental Demonstration of a Second-Law-Like Theorem Using a Colloidal Particle Held in an Optical Trap*, Phys. Rev. Lett. **92**, 140601, 1–4 (2004),
 <https://doi.org/10.1103/PhysRevLett.92.140601> or
 [http://rscweb.anu.edu.au/~sevick/groupwebpages/papers/PRL92_140601\(2004\).pdf](http://rscweb.anu.edu.au/~sevick/groupwebpages/papers/PRL92_140601(2004).pdf)
- [24] R. Brown: *XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies*, Philos. Mag. **4**, 161–173 (1828),
 <https://doi.org/10.1080/14786442808674769> or
 http://www.damtp.cam.ac.uk/user/gold/pdfs/teaching/old_literature/Brown1828.pdf
- [25] A. Einstein: *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Ann. Phys. (Leipzig) **17**, 549–560 (1905),
 <http://dx.doi.org/10.1002/andp.19053220806>
english translation:  <http://einsteinpapers.press.princeton.edu/vol2-trans/137>