

## Szilard's heat engine

M. O. MAGNASCO

*The Rockefeller University - 1230 York Avenue, New York, NY 10021, USA*

(received 2 December 1994; accepted in final form 25 January 1996)

PACS. 05.40+j - Fluctuation phenomena, random processes, and Brownian motion.

PACS. 05.70-a - Thermodynamics.

PACS. 87.10+e - General, theoretical, and mathematical biophysics (including logic of biosystems, quantum biology, and relevant aspects of thermodynamics, information theory, cybernetics, and bionics).

**Abstract.** - Szilard presented the first concrete embodiment of a Maxwell demon. We present a detailed kinematic analysis of his heat engine. We find that the phase space contains a branched manifold. After defining carefully the physical dynamics on such an object, we prove that the engine must operate at a loss.

In 1870, Maxwell [1] attempted to show the statistical character of the second law, through the device now called Maxwell's demon, which showed that manipulation of single molecules permits the creation of temperature gradients from an initially isothermal condition. Szilard [2] tried to understand how much energy the demon would consume in its operations, since Maxwell's description seemed to imply such consumption could be made arbitrarily small. Szilard's version consists of a movable partition in an isothermal cylinder with *one single* molecule as its operating gas (fig. 1). An observer determines on which side of the partition the molecule is in, attaches some weights to the appropriate pulley, and allows the "gas" to expand isothermally, doing  $kT \ln 2$  work on the weights. He then removes the partition, reinserts it in the middle, and starts all over again.

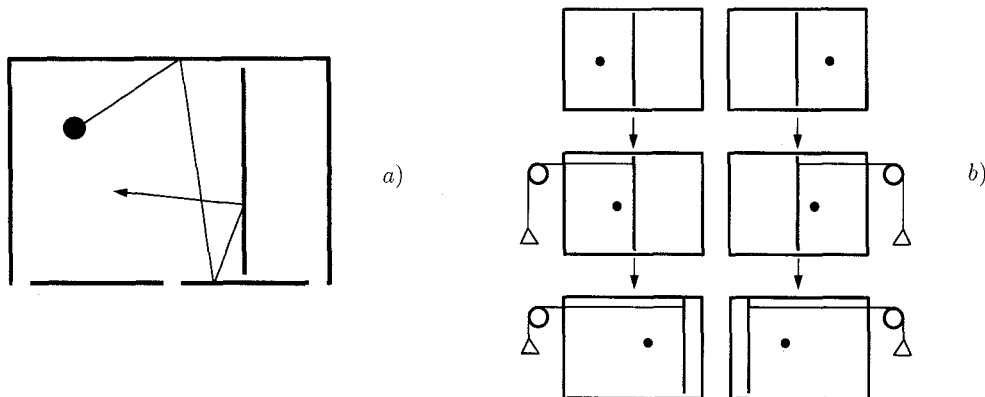


Fig. 1. - The operation of the Szilard heat engine. *a)* The one-molecule gas and the partition. Collisions are not ballistic: the particle exchanges energy with the cylinder and the partition. *b)* If we know on which side the molecule is, we know how to attach a weight.

Szilard did not prove this device not to work, but rather assumed it not to, and argued that the  $kT \ln 2$  gained in the expansion must be offset by an equivalent loss in order to preserve the second law; he attributes this loss to the measurement process. What Szilard meant by measurement was subsequently misinterpreted, and many people concentrated on the “gathering of information” process. Landauer [3] had the insight that the problem lay not in gathering, but actually in storing the gathered information. In particular, the demon needs to *erase* memory of previous measurements in order to make room for new ones. Resetting a “bit” of information from two possible values to just one is equivalent to compressing its phase space by a factor of two; this should require  $kT \ln 2$  of energy. Subsequently, Bennett [4] showed that measurement (in the sense of gathering) could be performed at zero cost.

However, it was argued that measurement is entirely unnecessary in the system [5]-[9]. The particle itself “knows” in which side it is, and can only push in one direction; the only device required is one where the weight is lifted as the partition is displaced from the midpoint (an absolute value function) [7], [9], [10]. The extraction and reinsertion of the partition can also be performed without “measuring” anything. Landauer [11], and Leff and Rex [12], claimed that some compression of phase space still takes place in resetting the wall to the center. Fahn presents an interesting heuristic argument [13].

The aim of this paper is to analyze with some rigour these ideas. Landauer’s argument is purely kinematic, so I will also do kinematics: the core of this paper is the analysis of the phase space of this engine. Since Landauer’s idea seems to be transparently correct, I should explain the need for further rigour. I will show that the phase space for the machine contains a strange topological object: a branched manifold. The problem is that normal thermodynamics does not hold on phase spaces which are sufficiently weird [14], a well-known problem in general relativity [15]. Thus the presence of a weird topological object in the engine’s phase space is a legitimate formal difficulty that has to be solved. I will show that the branched manifold, though weird, can be tamed: random motion through phase space will result in no net work being extracted from the engine, regardless of how pulleys or other contraptions are arranged.

I will first analyze the core of the phase space: the degrees of freedom associated to the particle and the partition. The wall’s position can be described through the two coordinates of its upper tip,  $x_w$  and  $y_w$  (fig. 2 a)). They are not independent: the partition can move along the cylinder (changing  $x_w$ ) only when fully inside ( $y_w = 1$ ) or fully outside ( $y_w = 0$ ). It can enter or leave the cylinder (changing  $y_w$ ) only through one of three doors ( $x_w = 0, \frac{1}{2}, \text{ or } 1$ ). Therefore, the phase space for the wall alone looks like a squarish figure 8 on its side, as shown

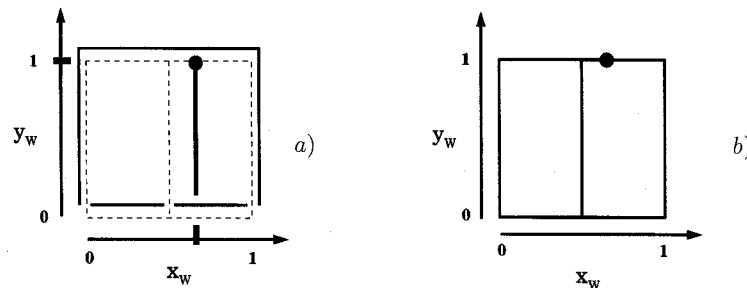


Fig. 2. – a) We call the coordinates of the upper tip of the wall  $x_w$  and  $y_w$ . The full range of motion of this upper tip is shown in dotted lines. b) The projection of phase space on the  $(x_w, y_w)$ -plane, which is identical with the “full range of motion” shown in dotted lines in a). The wall position shown in a) is shown as a thick dot. The wall’s coordinates (*i.e.* the dot) can move *continuously* along the lines of this graph: along the  $x_w$  direction while either fully in or fully out, or along the  $y_w$  direction when it is moving in or out of the cylinder through any of the three openings.

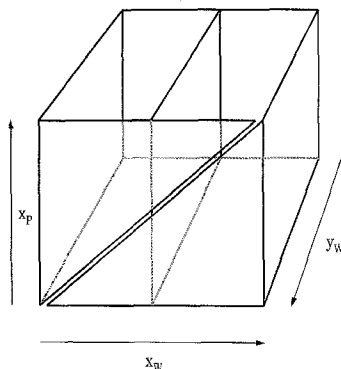


Fig. 3. – The full phase space of the system. The front face is the state when the partition is fully inside the cylinder ( $y_w = 1$ ). Gray represents “hidden lines”. The entire figure consists only of vertical surfaces. If viewed from the top, we recover fig. 2 b).

in fig. 2 b). Notice that this space is actually one-dimensional, in the sense that local motion is either forwards or backwards along a “track”; only the two joints at  $x_w = \frac{1}{2}$  are special.

Only the  $x$  coordinate of the particle,  $x_p$ , is relevant;  $x_p$  is kinematically independent of wall position while the wall is fully out, or partly through a door. However, when the wall is fully in, the particle and wall cannot cross each other, and hence the phase space for  $x_w$  and  $x_p$  while  $y_w = 1$  is a square with a diagonal slash removed (the front facet of fig. 3), forbidding wall and particle to cross each other.

Then, if we plot together all allowed values of the three variables in 3D space, we get the core of the phase space, shown in fig. 3. It is easier to describe how to make it than how it looks. Imagine gluing four identical squares to form the sides of a cubic box without top or bottom. Now glue another identical square, vertically, between the middle of two opposing faces; take one of these opposing faces, and cut it with a razor along a diagonal; the construction does not fall apart because it is held together by the middle wall, which is glued to *both* of the triangles left after the cut. Remember that even though we have to use three dimensions to build this object, it is locally two-dimensional: it is just a *surface*, it can be made by gluing paper strips.

This object is not that simple: it has cuts and seams. We can get a better look at it if we turn it inside-out, by making the inner wall go around the object through the outside in a wide circle, so that the two outer wings now lie next to each other, as in fig. 4. The object depicted is a “branched manifold”, and is called the Williams template [16] of the Bernoulli shift [17]. These objects have been employed extensively in dynamical-systems theory [18]. Their mathematical definition requires the seam to be tangential, so continuity of the tangent bundle is preserved.

The topology of this object poses a problem: how is a Brownian dynamics defined on it? *Prima facie*, it would seem that a random walker on the surface shown in fig. 4 could move perpetually forward, since it is forever decompressing. However, the trajectory will not only cross the seam from left to right, but also from right to left. If we naively let the walker cross undisturbed from left to right, but toss a coin to decide the branch upon going right to left, then the walker will indeed cycle through; detailed balance will be lost on this surface even in the absence of force fields, at constant temperature. But this is a physically incorrect definition of the dynamics. If we think of a fully viscous Brownian walker, then we must request it to be Markovian always. Thus coins have to be flipped every time the walker arrives at the seam, to decide which of the *three* branches it will take; and the odds for each should be independent of which branch it came from. Walkers with some amount of inertia pose a more

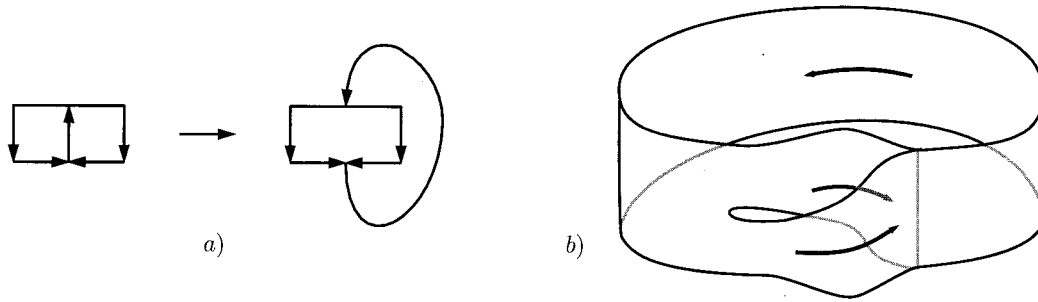


Fig. 4. - A transformation of fig. 3, to show it explicitly as a Williams template. *a)* We take the inner wall and move it to the outside. Arrows indicate the “intended” sense of operation, as in fig. 1 *b)*. *b)* After applying *a)* and smoothing some corners we get this representation. Since we have used the transformation in *a)*, the  $x_w$  and  $y_w$  axes are now mixed up, but the vertical direction is still  $x_p$ .

delicate problem: the speed is a tangent vector to the manifold, and there are three different possibilities as to how to glue the seam to preserve continuity of the tangent bundle. I will sidestep this problem entirely.

We can safely assume there is a bit of backlash, *i.e.* that the wall is slightly shorter than the height of the cylinder, *etc.* In that case the figure 8 of the wall’s phase space becomes a network of finite-width strips; the 3D phase space is depicted in fig. 5. Though it may look like an Escher perspective trick, it is not: it is a honest, solid piece of  $\mathbf{R}^3$ . It is something that could be contrived out of microwave waveguides, just a piece of *plumbing*, and if filled with gas, the gas will not spontaneously circulate around it. Nor would a random walker. The topological complication of the object has just vanished, it is only its boundary that retains some topology. But the boundary is unimportant, as I will show below.

Now we are ready to prove that the engine operates, necessarily, at a loss. We are not interested in what will happen to the engine at some particular period in time, but, rather, what will be the long-term outcome of its operation. So we want to consider the equations describing probabilistic (averaged) behaviour for a system with few degrees of freedom, in the presence of an isothermal bath. Both the particle and the wall are in contact with the bath and exchange energy with it; it is the randomizing (ergodic) effect of the thermal bath which allows us to replace the long-time average with a probabilistic spatial average. I will argue in the overdamped (Fokker-Planck) limit; the underdamped (Smoluchowskii-Klein) limit is just as trivial, only more laborious.

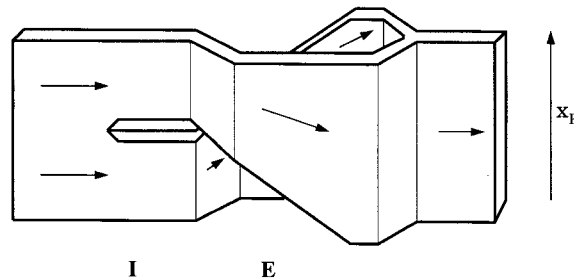


Fig. 5. - A “thick” version of fig. 4. The loop around and behind is omitted. Now the phase space is no longer just surfaces, but rather an object with finite thickness. Arrows and axes as in fig. 4 *b)*. I: insertion of the wall into the cylinder, E: gas expansion (two branches, as in fig. 1 *b)*).

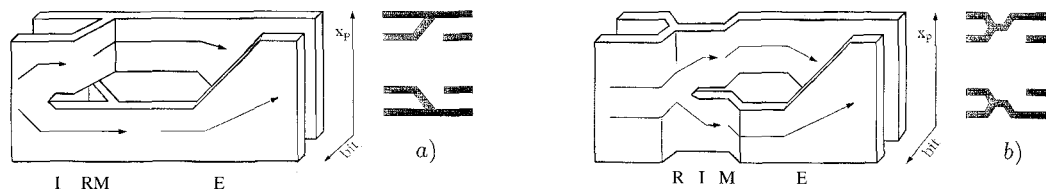


Fig. 6. – Phase space with an information bearing degree of freedom (away from the viewer; the two parallel tracks represent the two distinct bit states). I: wall insertion, R: bit resetting, M: measuring ( $x_p$ ), E: expansion. The gray polygons on the right show horizontal cross-sections of the top and bottom half, respectively. Notice that if motion goes from left to right (as intended in fig. 1), then the cross-sections reset a bit from an arbitrary state into a definite state which depends on  $x_p$ . a) The measurement and resetting take place simultaneously after insertion. b) The bit is “pre-reset” to a neutral intermediate state prior to wall insertion. This case is homeomorphic to fig. 5.

For an arbitrary mechanism of operation, we would like to include in our phase space all relevant degrees of freedom. We will deem all relevant degrees of freedom to be exhausted when the forces due to pulleys and weights are solely a function of position in phase space, rather than history or external intervention. We have thus a phase space (which will be some domain in  $\mathbf{R}^n$ ) where we have a vector field specifying the forces involved in making the machine work; we will call this vector field  $\mathbf{f}$ . Since we want to prove that the engine operates at a loss *regardless* of the arrangement of pulleys or other contraptions, and since  $\mathbf{f}$  represents this arrangement, we will have to keep  $\mathbf{f}$  arbitrary. I will not even need to request that it be curl-free.

The time-independent Fokker-Planck equation (in Euclidean space) says that  $\nabla \cdot \mathbf{J} = 0$ , where  $J$  is the probability current, given by  $\mathbf{J} = \mathbf{f}P - kT\nabla P$ , with  $P$  being the stationary probability density and  $\mathbf{f}$  as above.

The net power being consumed by the machine (the power spent in operating it minus the power obtained by lifting weights) is  $R = \int \mathbf{f} \cdot \mathbf{J} dV$ . If  $R < 0$ , then we are getting energy “for free”, *i.e.* cooling a single temperature bath. Because the case under consideration is *isothermal*, energy flows are directly proportional to entropy flows, so  $R$  is the rate of increase of entropy in the system. If we prove that  $R \geq 0$ , then we have both proven that the machine operates at a loss, and also the second law. From the definition of the Fokker-Planck equation,  $\mathbf{f} = \mathbf{J}/P + kT\nabla \log P$ , so that

$$R = \int \frac{\mathbf{J} \cdot \mathbf{J}}{P} dV + \int \nabla \cdot (\mathbf{J} \log P) dV. \quad (1)$$

The second term evidently vanishes, so that  $R \geq 0$  as expected. In fact,  $R = 0$  (we break even) only when the current vanishes identically and we have detailed balance. Please notice a *very* important point: this derivation *only* applies to a regular domain in Euclidean space. It does not apply on an object with a weird topology or geometry, because we need the divergence theorem to cancel the second term. So we see directly the role of taming the topology: the divergence theorem does not apply on a weird space, but it does apply on a regular subset of Euclidean space no matter how weird its boundary.

This fully finishes the analysis for any mode of operation of the machine, be there information-bearing degrees of freedom or not. However, for the sake of clarity, I will now present the case in which there is a single “bit” degree of freedom involved. Let us imagine the branched manifold again, but now we add an extra degree of freedom which will bear one bit of information; we can depict this degree of freedom as going in the *inwards* direction. There is no need to invoke backlash to regularize the topology since the extra dimension can do it by

itself. We represent the bit as an "ideal" bit, in that the two states are separated by an infinite barrier. In fig. 6 I depict the "bit" as towards the viewer. Notice, in fig. 6a), that right after resetting and measurement have taken place, the only available phase space is on the near track if  $x_p < \frac{1}{2}$ , or on the far track otherwise, which means that in this piece of phase space the state of the bit is in one-to-one correspondence with whether the particle is on the left or on the right. Figure 6b) is almost equivalent to 6a) (just a change of genus in the surface), and it is topologically identical to fig. 5. Hence, the argument in ref. [11] and [12], that even if no information-bearing degrees of freedom are present, measurement and resetting take place implicitly, is precisely correct.

\*\*\*

I would like to thank P. LAX, for first pointing out to me the paper by Szilard, and to G. CECCHI, L. FAUCHEUX, A. LIBCHABER and G. STOLOVITZKY for many discussions.

#### REFERENCES

- [1] MAXWELL J. C., *Theory of Heat* (Longmans, Green and Co., London) 1871.
- [2] SZILARD L., *Z. Phys.*, **53** (1929) 840-856.
- [3] LANDAUER R., *IBM J. Res. Dev.*, **5** (1961) 183-191.
- [4] BENNETT C. H., *Int. J. Theor. Phys.*, **21** (1982) 905-940.
- [5] POPPER K., *Brit. J. Phil. Sci.*, **8** (1957) 151-155.
- [6] FEYERABEND P. K., in FEYERABEND P. K. and MAXWELL G. *Mind, Matter and Method: Essays in Science and Philosophy in Honor of Herbert Feigl* (University of Minnesota Press, Minneapolis), (1966) pp. 409-412.
- [7] JAUCH J. M. and BÁRON J. G., *Helv. Phys. Acta*, **45** (1972) 220-232.
- [8] POPPER K., *The Philosophy of Karl Popper* (Open Court Publishing Co.) 1974, pp. 129-133.
- [9] ROTHSTEIN J., in LEVINE R. D. and TRIBUS M., *The Maximum Entropy Formalism* (MIT Press) (1979), pp. 423-468.
- [10] SVOBODA A., *Computing Mechanisms and Linkages, MIT Radiation Lab Series*, Vol. **27** (McGraw Hill) 1948.
- [11] LANDAUER R., *Phys. Scr.*, **35** (1987) 88-95.
- [12] Several of the above references, plus others and a lucid commentary are available in: LEFF H. S. and REX A. F., *Maxwell's Demon: Information, Entropy, Computing* (A. Hilger (Europe) and Princeton U.P. (USA)) 1990.
- [13] FAHN P., to be published in *Found. Phys.*
- [14] ZINN-JUSTIN J., *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford) 1989.
- [15] WALD R., *General Relativity* (University of Chicago Press) 1984.
- [16] BIRMAN J. S. and WILLIAMS R., *Topology*, **22** (1983) 47.
- [17] HOLMES P., in *New Directions in Dynamical Systems*, edited by T. BEDFORD and J. SWIFT (Cambridge University Press) 1988.
- [18] MINDLIN G. B., HOU X., SOLARI H. G., GILMORE R. and TUFILLARO N. B., *Phys. Rev. Lett.*, **64** (1990) 2350.