MAXWELL'S DEMON: THERMODYNAMICS OF INFORMATION GAINING AND INFORMATION PROCESSING

Aspasia Moue, Kyriakos Masavetas, Haido Karayianni

In 1867, James Clerk Maxwell proposed that a being, able to measure the microscopic parameters of a physical system and act accordingly, might be able to violate the Second Law of Thermodynamics. The being became known as Maxwell's demon. Information theoretic investigations conducted over the course of a century have placed well-defined thermodynamic bounds upon the action of Maxwell's demon. We argue that these attempts are ineffective, whether they follow Szilard in seeking a compensating entropy cost in information acquisition or Landauer in seeking that entropy cost in memory erasure. The logical irreversibility and missing information can derive limits upon the action of Maxwell's demon, but the understanding of the Maxwell's demon operation, as a device capable of measuring and manipulating physical systems, derives limits upon the application of information theoretic methods in Thermodynamics.

I. INTRODUCTION

The first thought experiment in which mention is made of information as a parameter with physical significance and linked with entropy is Maxwell's demon concept of a demon. Maxwell first introduced the demon in a letter to Tait on 1867 (Maxwell 1867). He imagined a being capable of differentiating molecules according to their speed. Maxwell's demon is a mechanism endowed with an intellect performing a rational function not requiring energy. This mechanism passes "hot" molecules into one half of the cylinder, and "cold" molecules in the other half by the opening and closing of a microscopic door in the wall dividing a gas cylinder into two halves. As a result of this sorting each half of the cylinder requires a different temperature $T_1 > T_{ex} > T_2$, where T_{ex} is the original temperature equal to the temperature of the external medium. In this way, a temperature difference between the two chambers is achieved without any work being done, contradicting the second law of Thermodynamics.

Since the invention of Maxwell's demon to this very day, the question about whether it is possible to design a perpetual motion machine that convert heat from the environment into work, has challenged some of the best scientific minds. The research on Maxwell's demon can be viewed carefully in terms of three major phases (the reader is referred over all this research, first to (Leff and Rex, 1990, 2003) for an excellent article's collection, second to (Moue et al, 2004) and references therein). The first phase covers the first 62 years of its existence (1867-1929), examines the possibility to construct mechanical (non-intelligent) demons and concludes that such systems could not operate because of the Brownian motion. The second phase covers the 'life' of Maxwell's demon from 1929 to 1961, examines the possibility to design intelligent demons and concludes that such systems are impossible because of the entropy cost in information acquisition (measurement). The third phase of the demon's life began on 1961, when R. Landauer made the important discovery that memory erasure in computers feeds entropy to the environment, and goes on up today. This discovery is now called 'Landauer's principle'. C. Bennett subsequently realized that Landauer's principle can be applied to a Maxwell's demon, who gathers information and stores it in its memory. A complete thermodynamic analysis of a demon's cycle operations requires that its memory be brought back to its initial state. The entropy sent to the environment by this erasure turns out to be *just large enough* to save the second law.

The study of Maxwell's demon activity threw new light on the thermodynamical significance of gaining information and the thermodynamics of information processing. In his seminal article (Szilard 1929), L. Szilard considers a device and states his objective as follows: "The objective of the investigation is to find the conditions which apparently allow the construction of a perpetual motion machine of the second kind, if one permits an intelligent being to intervene in a thermodynamical system". The appropriate way to think about this question is to consider whether a mechanical demon incorporating a computer could work as a perpetual motion machine – that is, a device with information-gathering and information-processing abilities.

The purpose of the present work is (i) to examine those information theoretic methods that have effected limits on the operation of Maxwell's demon and (ii) to show that the understanding of Maxwell's demon illustrates the possibility for the application information theoretic methods in Thermodynamics *but* not without clarifications and

PHYSICAL CHEMISTRY LABORATORY, DEPARTMENT OF CHEMICAL ENGINEERING, NATIONAL TECHNICAL UNIVERSITY OF ATHENS, 9 HEROON POLYTECHNIOY STR., ZOGRAFOU CAMPUS, 15780, ATHENS

limitations. In order to explain this purpose (in sections 3 and 4), the pertinent results of thermodynamics of computation are derived in the second section.

II. The thermodynamics of Computation

At first we review some relevant observations in the thermodynamics of computation, following the discussion in (Feynmann 1996).

The fundamental principle in the thermodynamics of computation is that information should be conceived as physically embodied in the state of a physical system. So we can, for example, think of a message on a tape – a sequence of 0's and 1's – as represented by a sequence of boxes, in each of which there is a 1-molecule gas, where the molecule can be either in the left of the box (representing the state 0) or the right half of the box (representing the state 1).

If we assume that the tape (the sequence of boxes) is immersed in a heat bath at constant temperature T, the amount of work, W, required to compress the gas isothermally in one of the boxes to half the original volume V is:

$$W=-kT\log 2 \qquad (1)$$

where k is Boltzman's constant.
So the entropy of the 1-molecule gas changes in this thermodynamically reversible change of state by an amount

$$\Delta S=-k\log 2 \qquad (2)$$

and the entropy of the environment is increased by $k\log 2$. Equivalently, there is a change of $kT\log 2$ in the free energy (F) of the gas (because in an isothermal compression in $\Delta F=-T\Delta S$).

The information in a message can be defined as proportional to the amount of free energy required to reset the entire message tape to zero, in the sense that each cell of the tape – each 1-molecule gas in the box – is compressed to half its volume, reducing the number of available microstates by half. In appropriate units (taking logarithms to the base 2), it takes 1 bit of free energy to reset each cell to a zero value.

R. Landauer introduced the concept of 'logical irreversibility' in connection with information-discarding processes in computers (Landauer 1961). Memory erasure which takes a computer memory from an (arbitrary) existing state A, to a unique, standard reference state L discards information in a logically irreversible way. Logical irreversibility means that the prescription 'Map the existing state A to the state L' has no unique inverse because state A can be any of many possible states in the computer's memory. The thermodynamic property in question is *dissipativity*. A thermodynamic process is dissipative if it is entropy increasing, that is, if it takes energy in a form usable as work and turns it into useless heat energy. That logically irreversible computation is dissipative has been the prevalent opinion ever since Landauer proposed it.

According to Landauer and for a little more detailed discussion consider Fig. 1, in which is shown a typical logical process, which discards information, e.g., a logical variable that is reset to 0, regardless of its initial state. Fig. 1.1 (based on Landauer 1992) shows, symbolically, the phase space of the complete computer considered as a closed system, with its own power source. The erasure process we are considering must map the 1 space down into the 0 space. Now, in a closed conservative system phase space cannot be compressed, hence the reduction in the [horizontal] spread must be compensated by a [vertical] phase space expansion, i.e., a heating of the [vertical] irrelevant degrees of freedom, typically thermal lattice vibrations. We are involved here in a process which is similar to adiabatic magnetization, i.e., $k\ln 2$ per erasure process. Fig. 1.2 shows the end-result of the erasure process in which the original 1 and 0 spaces have both been mapped into the [horizontal] range originally occupied by the 0. This is, however, rather like the isothermal compression of a gas by the cylinder into half its original volume. The entropy of the gas has been reduced and the surroundings have been heated, but the process is not irreversible, the gas can be subsequently expanded again. Similarly, as long as 1 and 0 occupy distinct phase space regions, as shown in Fig. 1.2, the mapping is reversible. The real irreversibility comes from the fact that the 1 and 0 spaces will subsequently be treated alike and will eventually diffuse into each other.

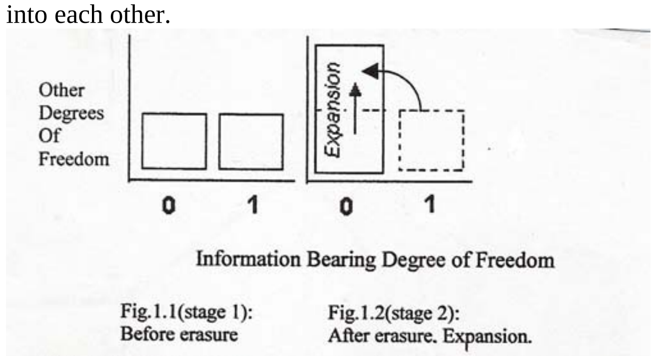

Fig.1. Memory erasure according to Landauer

Landauer also argued that computation steps that do not discard information, e.g., writing and reading, can be thermodynamically reversible in principle (Landauer 1992, Bennett and Landauer 1985). C. Bennett extended Landauer's work, arguing that a computing automaton can be made logically reversible in any step (Bennett 1982). This allows an in-principle thermodynamically reversible computer that saves all intermediate results, avoiding irreversible erasure, prints out the desired output, and reversibly disposes of all undesired intermediate results by retracing the program's steps in reverse order, restoring the machine to its original condition. These results are known as "*Theory of reversible computation*".

In this way and under the assumption that the essential feature of a measurement is that it establishes a correlation between the state of a system and the state of a memory register, the establishment of a correlation between the states of the two systems is equivalent to a copying operation, and there is no entropy cost to copying (therefore to measurement too).

III. THE INFORMATION THEORETIC LIMITS ON MAXWELL'S DEMON

Broadly speaking, there are two approaches.

The *first* approach was presented most clearly in Szilard's famous paper (Szilard 1929) and later, was embodied in Brillouin's (Brillouin 1962, Chapter 13), Zurek's (Zurek 1984, pp. 151-161), Fahn's (Fahn 1996) and Leff and Rex (Leff and Rex 1994) work. This first approach sees an entropy cost in information acquisition. Its basic postulate is that gaining information that allows us to discern between n equally likely states must be associated with a minimum entropy cost of $k\log n$, the entropy dissipated by the system that gains the information.

Szilard's model allows thermodynamic analysis and interpretation, but at the same time entails a binary decision process. Thus, long before the existence of modern information theory, Szilard focused attention on the 'information' associated with a binary process and connected the Maxwell's demon puzzle with *information*.

A Szilard's version of demon is illustrated in Fig. 2.

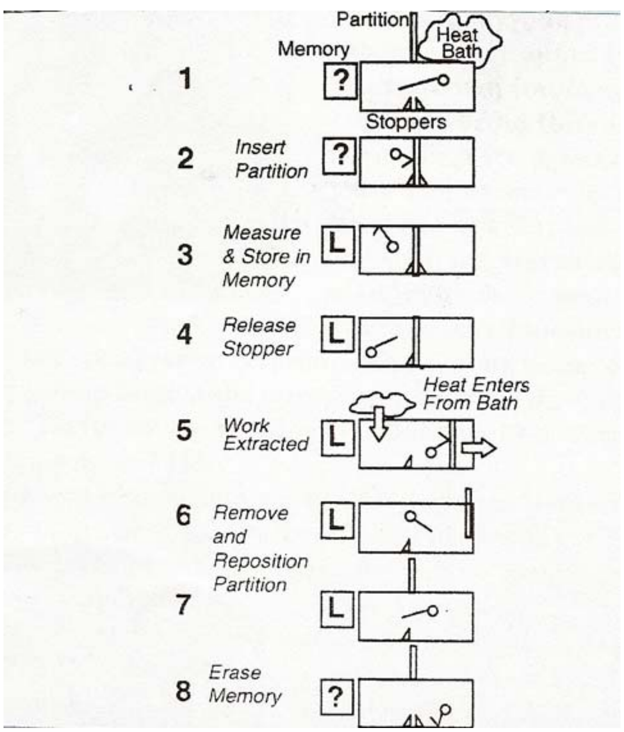

Fig. 2. Szilard's version of Maxwell's demon.

At stage 1 a gas consisting of a single molecule is in thermal equilibrium with a heat bath. At stage 2 a partition is inserted through a slit in the container wall, and held in place by stoppers. Assume, for simplicity, that the partition divides the container in half. At stage 3 a 'demon' measures whether the particle is on the right or the left hand side of the partition and stores the information in its memory (only the demon's memory appears in the illustration). At stage 4 the demon releases one stopper according to the contents of its memory, allowing the partition to move towards the empty side. The partition becomes a piston that is pushed by the pressure of the gas. At stage 5 the pressure of the ideal gas produces an amount of work $\int p\mathrm{d}V = kT\ln 2$ that is stored in a work reservoir. At stages 6 and 7 the partition is removed and restored to its original position, and at stage 8 the demon's memory is erased. The operation cycle is thereby closed, and the only change is a change of entropy: heat energy has turned into energy available as work.

Szilard's idea was that the demon produces work (at stage 5 of Fig. 2), but *according to the second law of Thermodynamics*, one of the stages necessary for this production (stage 3 of the measurement) is dissipative, by an amount that compensates for the entropy reduction. This dissipation solves the problem, since the net entropy balance becomes zero or positive.

The *second* approach was presented by Bennett (Bennett 1987) through a device that has been called 'Szilard's engine' by Bennett, since it is a similar device of Szilard's 1-molecule heat engine. A series of articles by Bennett (Bennett 1988, 1998), Landauer (Landauer 1997, 1996a, 1996b), Shizume (Shizume 1995) and Piechocinska (Piechocinska 2000) was illustrated the association of the Maxwell's demon puzzle with *computation*. In Szilard's engine, the demon can insert a partition that separates the box into two equal parts, left (L) and right (R). Initially, the demon's memory register is in a neutral or ready state, 0. Bennett's idea was that the demon produces work (at stage 5 of Fig. 2) and achieves an entropy decrease of 1 bit. However, *by the second law*, the entropy cannot decrease and *because of this reason*, Landauer and Bennett argue that this entropy decrease of 1 bit in the system must be accompanied by an entropy increase in the environment of 1 bit, that is, a minimum entropy of 1 bit must be dissipated to the environment in erasing and resetting the memory to 0 (at stage 8 of Fig. 2).

The claim that information erasure, and not information acquisition, involves a minimum entropy cost depends on the assumptions that (i) measurement is essentially a copying operation with no entropy cost in principle, (ii) reversible computation is possible in principle, and (iii) erasure involves compressing the phase space of the physical system that functions as a memory register, which requires dumping heat into the environment. The last assumption ((iiii)) is the 'Landauer's principle' accordingly to which in erasing information that can discern n states, we dissipate at minimum entropy $k\log n$.

Bennett's and Landauer's approach is currently accepted for the most researchers, as the final work on the threat of Maxwell's demon. Investigators have now turned toward understanding the demon.

IV. THE UNDERSTANDING OF MAXWELL'S DEMON LIMITATIONS ON INFORMATION METHODS IN THERMODYNAMICS.

While the basic assumptions in the above arguments of Szilard, Bennett and their followers seem reasonable and create fundamental elements in the connection between Thermodynamics, information theory and computation, we believe that the understanding of Maxwell's demon as a case study arises limitations on this connection. Maxwell's demon consists a case study of this connection because it is a system that is both informational and thermodynamic. It is *informational* because of its information-gathering and information-processing abilities. Also, it is *thermodynamic* because it works as a thermal engine.

The most unsound idea in both kinds of argument (viz. the argument which establishes that measurement is dissipative and the argument which establishes that memory erasure is dissipative) is the fact that both arguments *rely on* the second law, as the key to the eliminative process (see p. 3 of the present work). Whereas relying on the second law is normally very plausible, this is not the case where Maxwell's demon is concerned, since the demon is meant to serve as a counterexample for this law. We need an independent argument for the dissipation in the operation of the demon, one that relies neither on the demon nor on the second law. The arguments in the previous section claim that the demon demonstrates alleged principles in the thermodynamics of information, regarding the entropy of measurement and memory erasure. These provide the dissipation that compensates for the demon's work. Presently, however, there is no available satisfactory independent justification for any of these ideas, and present technology cannot test them empirically.

Furthermore, in these arguments there are, also, the following unsound ideas:

- The idea of 'Szilard's school' in its 'informational machine' that a single molecule can be a 'gas' is problematic. Szilard needs it in order to use the ideal gas law. But Szilard's version, through which the connection between information and entropy started, is inadequate for investigating the second law because a single molecule means that the insertion of a partition (stage 2 of Fig. 2) confines the gas to half of the container without investing work, thereby violating the Gay-Lussac's law.
- The investigation that 'Szilard's school' give for the perpetual motion machine of the second kind has to do with the long run, average result of many repeated cycles of operation. We cannot exclude the possibility of the second law being violated by any simple operation cycle.
- The work of 'Landauer's school' on reversible computation, on which is based its argument about a 'paradox' of Thermodynamics as Maxwell's demon is, is illuminating, but it is important to realize that this work does not constitute formal Thermodynamics. The problem lies in the fact that not only one single broadly applicable thermodynamic equation is presented in any of the work on reversible computation. The only mathematics in that body of work either involves abstract computation theory (for example, Turing machine theory) or idealized situations by means illustrative engineering examples. In particular, the crucial conjecture that entropy must increase in any many-to-one mapping is nowhere formally proven. In fact no broadly applicable sequence of equations involving the Gibbs canonical fine-grained entropy $-k\int \mathrm{p}(X)\ln \mathrm{p}(X)\mathrm{d}X$, Gibbs fine grained entropy $k\ln V_{(a.r)}$ or Boltzmann entropy $k\ln V_{(mac)}$ is presented anywhere. Here, in the first expression p is the probability distribution function, X stands for all the degrees of freedom and the integration is taken over all phase space. In the second expression $V_{(a.r)}$ is the volume of the accessible region and in the third expression $V_{(mac)}$ is the phase space volume of a *macrostate*.

  This lack of an overaching theoretical framework forces the field of reversible computation to rely heavily on reasonableness arguments and on generalizing from particular examples. Unfortunately, for the case of Maxwell's demon, neither such kind of argument can *prove* anything. Moreover, these kinds of arguments are extremely limited in scope, as for example, the usual argument that total entropy must increase in any many-to-one mapping, whether that mapping occurs in a computer. However, as has been pointed out by J. Berger (Berger 1990) among others, this argument does not establish that total entropy *must* increase in a many-to-one mapping, only that such an increase is reasonable.
- Landauer, although he provides little explanation, seems to be applying methods of statistical thermodynamics, not at the molecular level but to a system comprised of N *macroscopic* subsystems, the bits. Not only is this procedure questionable, but the process considered has no thermodynamic significance.
- The system of Landauer in Fig. 1, which depicts the phase space of the whole computer, is subject to Liouville's theorem. That is a central assumption in the Landauer's argument, for it is Liouville's theorem that dictates the constancy of phase volume, forbidding the reduction in phase space volume in the transformation of Fig. 1.1 and dictating its direct transformation to Fig. 1.2. Also, by Liouville's theorem both Gibbsian expressions of entropy are constant in time. Since the phase volumes of Fig. 1.1 and 1.2 are equal throughout the transformation (with uniform probability distribution inside and zero outside), their Gibbs microcanonical (as well as canonical) entropies in them are equal. *No Gibbsian dissipation takes place* at the transformation from Fig. 1.1 to 1.2. Hence, the dissipation that Landauer proclaims cannot be of *this* sort.

In Boltzmann's approach, entropy can change spontaneously, when the system evolves from on macrostate to another. Whereas Gibb's entropy is determined by the volume of the accessible region, and hence this volume must be carefully identified, Boltzmann's entropy is determined by the volume of the macrostate, and so the macrostates have to be carefully identified. Above is mentioned that the memory states 1 and 0, in the Landauer's argument, are the natural macrostates of a memory cell: distinction between them is sufficient for manipulating bits. Since this is an important point, a possible objection ought to be addressed: In the computers we currently use individual memory cells are certainly inaccessible for the normal user of any practical level. In what sense, then, are the1's and 0's, or arrays of them, distinguishable macrostates? The input and output of a computer that performs an algorithm are very well defined arrays of bits. A flip of a bit may change the input and output in a way that is very significant at the pragmatic level. Therefore the user must be able to distinguish 1 from 0 at both ends. It makes no difference whether the distinction can be directly perceived with naked human senses or using complex and theory-dependent devices. The use of automated error correction processes should make no difference either. The point here is only the very distinguishability: if and only if 1 and o can *somehow* be distinguished, they are different macrostates. And if they are distinguishable at the input stage there is no reason why they should cease to be so during computation. And so, prior to the erasure (Fig. 1) the Boltzmann entropy is either $k\ln V(0)$ or $k\ln V(1)$, and after the erasure it is $k\ln V(0)$. In the simplest or extreme case $V(0)=V(1)$, and the entropy difference is null. In other words, entropy is the same before and after erasure. *No Boltzmannian dissipation takes place*. We conclude that the 'Landauer's school' argument, even if it were correct (supplemented, for example, by the idea of algorithmic complexity), couldn't have solved the *thermodynamic* 'paradox' of demon, because the meaning of *dissipation* in this argument is not connected to the usual notion used in Thermodynamics and statistical mechanics. Therefore, whether or not erasure is dissipative for the information theory is *irrelevant* for the demon.

- The idea of application of 'Landauer's principle' in Maxwell's demon problem is questionable because has been shown that even logically irreversible operations of a computerized demon (as the memory erasure) can in principle be accomplished in a thermodynamically reversible fashion without the need to erase information and entropy cost (Moue et al 2005).

V. Conclusions

More than a century after Maxwell introduced the idea of an information gathering and system manipulating entity, Maxwell's demon remains an intriguing point of investigation. By its ability to act and compute in the microscopic domain, the demon involves information theoretic investigations to problems in Thermodynamics for which statistical methods fail. On one hand, the logical irreversibility of erasure puts limits on the demon's interactions with systems, On the other hand, the understanding of the demon's operations puts limits on the connection between Thermodynamics and information theory and, as a result, the demon seems an excellent source for both practical and theoretical investigations of information theoretic methods in Thermodynamics.

VI. References

[1] Bennett, C. H. (1982): The Thermodynamics of Computation – a Review, *International Journal of Theoretical Physics*, **21**, pp. 905-940.

[2] Bennett, C. H. and Landauer, R. (1985): The Fundamental Physical Limits of Computation, *Scientific American*, **253**, pp. 48-56.

[3] Bennett, C. H. (1987): Demons, Engines and the Second Law, *Scientific American*, **257**, pp. 108-116.

[4] Bennett, C. H. (1988): Notes on the History of Reversible Computation, *IBM Journal of Research and Development*, **32**, pp. 16-23.

[5] Bennett, C. H. (1998): Information Physics in Cartoons, *Super lattices and Microstructures*, **23**, pp. 367-372.

[6] Berger, J. (1990): Szilard's Demon Revisited, *International Journal of Theoretical Physics*, **29**, pp. 985-995.

[7] Brillouin, L. (1962): *Science and Information Theory*, New York: Academic Press Inc.

[8] Fahn, P. N. (1996): Maxwell's Demon and the Entropy Cost of Information, *Foundations of Physics*, **26**, pp. 71-93.

[9] Feynmann, R. P. (1996): *Feynmann Lectures on Computation*, Reading: Addison – Wesley

[10] Landauer, R. (1961): Irreversibility and Heat Generation in the Computing Process, *IBM Journal of Research and Development*, **5**, pp. 183-191.

[11] Landauer, R. (1992): Information is Physical, Proceedings of the Phys Comp 1992, IEEE Computer Society Press.

[12] Landauer, R. (1996a): The Physical Nature of Information, *Physics Letters* A, **217**, pp. 188-193.

[13] Landauer, R. (1996b): Minimal Energy Requirements in Communication, *Science*, **272**, pp. 1914-1918.

[14] Landauer, R. (1997): A Fundamental Physical View, *Physics Scripta*, **35**, pp. 88-95.

[15] Leff, H. S. and Rex, A. F. (1990): *Maxwell's Demon: Entropy, Information, Computing*, Bristol: Adam Hilger

[16] Leff, H. S. and Rex, A. F. (1994): Entropy of Measurement and Erasure: Szilard's Membrane Model Revisited, *American Journal of Physics*, **63**, pp. 994-1000.

[17] Leff, H. S. and Rex, A. F. (2003): *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, Bristol and Philadelphia: Institute of Physics Publishing.

[18] Maxwell, J. C. (1867): Letter to P. G. Tait, 11 December 1867, published in C. G. Knott (1911): *Life and Scientific Work of P. G. Tait*, London: Cambridge University Press, pp. 213-214.

[19] Moue, A., Masavetas, K. and Karayianni, H. (2004): Understanding the Development of the Concept of Entropy through the Context of the Maxwell's Demon Operation, *Chemical Educator*, **9**, pp. 204-210.

[20] Moue, A., Masavetas, K. and Karayianni, H. (2005): Maxwell's Demon, Thermodynamics and Information Theory, *IASME (International Association of Mechanical Engineers) Transactions*, **2** (1), pp. 59-67.

[21] Piechocinsta, B. (2000): Information Erasure, *Physical Review* A, **61**, pp. 1-9.

[22] Shizume, K. (1995): Heat Generation Required by Erasure, *Physical Review* E, **52**, pp. 3495-3499.

[23] Szilard, L. (1929): Über die Entropieverminderung in einem Thermodynamischen System bei Eingriffen Intelligenter Wesen, *Zeitschrift für Physik*, **53**, pp. 840-856.

[24] Zurek, W. H. (1984): Maxwell's Demon, Szilard's Engine and Quantum Measurements, in G. T. Moore and M. O. Scully (eds): *Frontiers of Nonequilibrium Statistical Physics*, New York: Plenum Press, pp. 151-160.