Applying Physics Methods to Information Theory

H.Grassmann,
University of Udine, Italy
and Isomorph srl.*

Currently, information theory is not a part of physics, and makes use of variables, which are not physics variables. We explore the possibility to apply known physics procedures to problems of information theory. We do this by considering whether some well defined mathematical structure can be associated to messages and message processing systems, and by defining variables which can be measured by means of well defined measurement procedures.

STATE OF THE ART

Information theory dates back to Shannon, who created it in order to study problems of signal transmission [1]. Shannon's major concern was the effect of noise, and for this reason his theory takes a probabilistic character. For instance, the amount of information of a message is defined in the following generally known way:
>>*If in a set of messages the probabilities of the possible messages are given by p_1, p_2, … p_n, then the amount of information associated with the first message is $log_2(1/p1)$, that of the second $log_2(1/p_2)$, and so forth. The expected value of these amounts of information is called the entropy H, or the average information of the message set.*<< (here we quote from Encyclopedia Britannica. For digital messages, and if 1 and 0 have the same probability to occur, the entropy of the message is therefore equal to the length of the message in bits.)
It is often assumed, that the information-"entropy" is identical to the physical entropy variable of thermodynamics. But this is not true. The entropy of physics is a state function, the "entropy" of information theory is not. For instance, a newspaper does not change its amount of information, when put into a refrigerator, while instead its thermodynamic entropy changes.
After Shannon, Chaitin, Kolmogorov and others created the algorithmic information theory, which is primarily concerned with message producing systems [2,3,4]. (While Shannon had gone deeper in studying the properties of messages.) The logical structure of algorithmic information theory is very similar to Shannon's information theory. In algorithmic information theory the complexity of a program is the minimum length of a program which is capable to produce a certain message. Again, this complexity is not a physics variable, since it is in general not clear what "the shortest program" is supposed to be.
Szilard had attempted to insert the concept of "information" in physics [5]. Studying the problem of the Maxwell demon, he concluded that storing one bit of information must necessarily dissipate the energy 1/2·k·T (k = Boltzmann constant, T = absolute temperature). Later, Landauer argued that not storing of information dissipates energy, but rather erasing [6]. Bennett instead concluded that in certain situations (reversible computing) energy dissipation in information processing can be avoided altogether [7].

* Prepared for the proceedings of the Rencontres de Physique de la Vallee d'Aoste, La Thuile, Aosta Valley, Italy, 24 February - 01 March 2008.

This situation – with the findings of Szilard, Landauer and Bennett being in part in agreement, in part not – is not satisfying, and it does not yet constitute a conclusive physics theory.

PURPOSE OF THIS PAPER

As far as thermodynamics is concerned, Szilard's arguments are based on a solid mathematical framework (statistics), provided by Boltzmann and others. However, Szilard applied thermodynamic arguments to the new field of "information", where there was no known mathematical structure at all. A possible mathematical structure of messages or message processing system was not known and was not considered. As a consequence concepts like "information" or "erasure" or "storing" were not defined in a general and fundamental way. Terms like "information" or "erasure" were clear in the context of a specific example (like Maxwell Demon or reversible computing etc), but had no general validity. Therefore statements, which were correct in a certain context where not necessarily correct in a different one.

In this paper we shall attempt to better understand the mathematical structure of messages and of message processing systems. This might help to apply Szilards thermodynamic arguments to problems of information processing, making thereby the physics of information a new field of physics research.

Since our contribution to the conference of La Thuile was given in the section "Science and Society" we also want to comment on the implications which the physics of information might have on fields like research on brain mode of operation, philosophy, theology and law, and in general on the position of the human being in this world.

MATHEMATICAL STRUCTURE OF MESSAGES AND OF MESSAGE PROCESSING

We briefly summarize the findings presented in [8]:

"1" and "0" are the elements of the remainder class modulo 2 [9], which is a field of dimension 2 commonly referred to as F_2. Over F_2 one can construct the arithmetic vector space F_2^n. Therefore digital messages are vectors in a mathematical sense. For instance, (1,0,1,1,0) is a 5-digit digital message, but it is also a element of F_2^5. It is therefore justified to discuss problems of information processing in terms of algebra. One can, for instance, consider the processing of an input message, *a*, by an information processing system, *T*, as a vector transformation, resulting in the output vector *b*: $b=T(a)$.

Considering information processing as an instance of vector operation allows us to connect mathematical concepts to concepts of the physical world by means of postulates. For instance, we can define when two messages have the same meaning: Some contract may have been translated from English into Chinese, and we would like to be sure that the Chinese version "is identical" or "has the same meaning" as the English original. We can verify this by having the Chinese version translated back to English – if the original and the re-translated text are the same, we know that also the Chinese version is correctly translated.

We can express this now in the more precise mathematical language as follows. We postulate that two messages, *a* and *b*, have the same meaning, and contain the same information with respect to an information processing system, *T*, if $b=T(a)$ and $a=T^{-1}(b)$. In

this way also the “information” itself gets defined: An injective System T will conserve information, or: information is, what is conserved by injective systems.
It seems furthermore reasonable to assume that two messages which contain the same information must have the same amount of information. But *T* being injective (*T* does not delete information) does not necessarily mean that *a* and *b* must have the same length (in number of bits). Therefore, in our scenario, the length of a message is not a measure of the amount of information of the message, in contrast to classic information theory.
In our scenario one might measure the amount of information of a message as the logarithm of the number of possible messages in a vector space, or the number of dimensions of the vector space.
For instance, if an information processing system outputs messages of n bit length, and if there are m different such messages, the amount of information of each message would be $\log_2(m)$, which becomes identical to the Shannon measure only for the case of $m=2^n$.
One can conclude that considering the vector nature of message allows us to connect the mathematical concepts of information processing to the physics world. This is a significant progress with respect to classic information theory where one cannot compare messages in any way, and where even the concept of “information” is not defined.
A digital message of n bit length can have 2^n different values. We assume, that *T* can process the message regardless of what value it has, that means, that all 2^n possible vectors are possible input messages. Therefore the input vectors form a vector space, we refer to it as *U*. We next want to investigate what properties the output vectors have, assuming certain properties of *T*. Throughout this note we always assume *T* to be injective.

LINEARITY OF INFORMATION PROCESSING SYSTEMS

If T is linear, it follows that also the output messages form a vector space, V, and input and output vector space are isomorphic. It also follows that T can be represented by a matrix with n columns, which span a vector space, which is isomorphic to both U and V. In different words: the vectors, which form T, can be injected to both U and V. This means that at execution time the information processing system does not input any new information. It rather consists of the information to be processed, already before run time.
Szilards argument (storing information must dissipate energy) can only be applied to the creation of T itself, since when creating T (loading of a program, or construction of a hardware device) some information obviously must be imported. Szillards argument does not apply to the operation of T , since T does not input new information while operating.
One may address at this point Landauers and Bennetts “counter examples” to Szillards findings [6,7] in some detail. Here we note simply that those examples were possibly based on the assumption, that when inputting and processing a message an information processing system necessarily inputs information (this assumption is quite natural in conventional information theory).
As far as Landauers and Bennets findings on the erasure of information are concerned [6,7] , we will later show that messages can be erased without energy dissipation.
Also if T is not linear there can be no more than 2^n different output values of the n-digit input message a. The 2^n different input values can be mapped onto a base with 2^n dimensions, and the same can be done with the output elements.
Then T can be replaced by an operation D operating between these two bases. Therefore D can be extended into a linear operation.

Said differently: T always can be executed by means of a look up table with 2^n elements. A lookup table is a special kind of a matrix.

We can conclude with two findings:
(1) the “linearity” or “non linearity” of an instance of physical information processing (we consider n to be finite) is not a fundamental notion, but rather refers to the technical particularity of the information processing system – e.g. the use of digital computers instead of canonical computers.
(2) For all information processing systems, also if they are not linear when operating in a digital space it is true: the information processing system must contain the information to be processed already before run time, and input and output messages have the same information.

In our scenario, the processing of a message is just a particular instant of message transmission. There is no explicit need for energy dissipation. And there is no need for a time sequence in information processing, information processing can occur on the light cone of the message. We want to illustrate these features by a *gedanken experiment* in the following chapter.

PHYSICAL COMPLEXITY

Figure 1 shows a very simple information processing system. It has a digital number as input and it multiplies this number by 2.

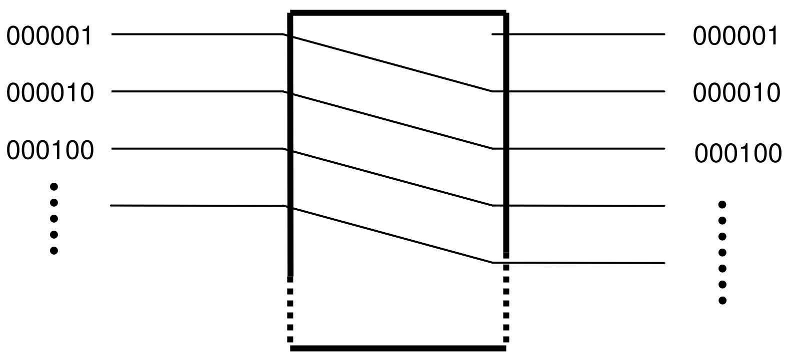

Figure 1: Device performing a multiplication of a digital input number by 2, y=2·x. The input signal arrives with the signal lines at the left of the device.

This operation occurs without energy dissipation and on the light cone of the message. The processing of the message and the transmission of the message cannot really by distinguished – the box, which indicates the “information processing system” in figure 1 could as well be omitted. It serves only to guide the eye, but it does not have any relevant function.
It is furthermore true, that there is no program operating and therefore one cannot determine the complexity of the program which performs the operation of multiplication.
With the same method we can perform any kind of operation if we operate with canonical numbers, instead of digital numbers. An example is shown in figure 2: the “box” of figure 2 has canonical numbers as input and output and it performs the operation $y=x^2$.

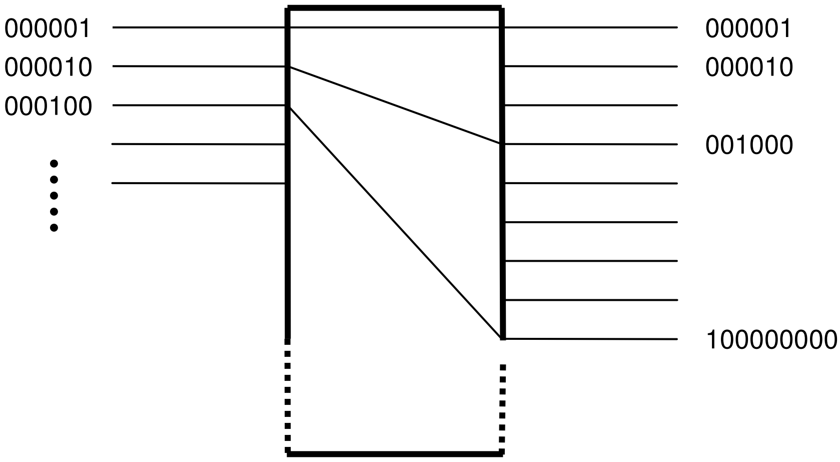

Figure 2: Device performing the operation $y=x^2$ on canonical numbers.

Again, one cannot talk about a "program" or its complexity.
The information processing boxes of figure 1 or 2 can also erase messages. For example one could connect all input signal lines to always the same output line, for instance one might choose the line representing "0". Regardless of what the input message is, the device would always produce the same 0 output and it would therefore have erased the message. This again would be done without energy dissipation (in contrast to the findings of Bennett).
The functionality of the computing boxes in figures 1 and 2 can be represented by a lookup table: as the devices in figure 1 and 2, a lookup table can produce for each input number a corresponding output.
In the most general case, in order to process an input number of n bits, the lookup table would need to have 2^n elements. But for many operations, a system of smaller lookup tables would be sufficient in practice for performing the operation. This is shown in figure 3: the device of figure 3 performs an addition of two numbers of two bits.
Each of the boxes indicated with an "+" adds two bits. These "fundamental adders" can be considered lookup tables themselves – they attribute to each of the 2^2 possible combinations of "1" and "0" the corresponding result.
In general, if we want to perform the addition of two numbers of n bits each by a system of lookup tables as in figure 3, we need $2 \cdot n$ fundamental adders, each with 4 memory places, for a total of $8 \cdot n$ lookup table elements. The complete lookup table would have instead 2^{2n} elements.
This suggests to define the complexity of an information processing system as the minimum number of lookup table elements needed for the operation, or as the logarithm thereof. The lookup table elements are physical entities, which can be counted without ambiguity. Therefore this "physical complexity" can be measured (in contrast to the complexity of algorithmic information theory). Once we have decided to use for instance digital numbers, the complexity of an operation is then well defined and quantifiable and it does not depend on the program language.
Still, in lookup table systems like in figure 3 the processing can be performed without the need of organizing the operation in time steps.
If we introduce a temporal sequence instead, we then can perform the whole addition with only one fundamental adder. We can do this by using again and again the same one lookup table (for adding two bits), and storing the intermediate results. This presents than

what one would refer to as a “program”. The complexity of the program might therefore be measured by the number of repeated uses of a lookup table (or the logarithm thereof).

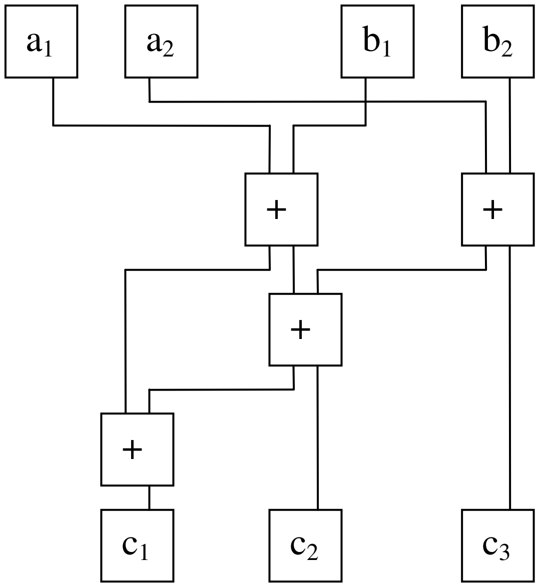

Figure 3: Machine for performing the calculation *a+b=c* on two input numbers of two bits each.

We have found in this chapter that one can perform mathematical operations like addition or multiplication with a system of lookup tables, which is smaller than the complete lookup table of 2^n elements (for n bits of input), without making any approximation. In the next chapter we will investigate, whether also an image recognition system can be built with a small size lookup table.

COMPUTING THE DESCRIPTION OF IMAGES

Without limiting generality, we assume an image of n black and white pixels (“1” and “0”). A universal image recognition system would consist of a lookup table of 2^n elements, each element would describe the corresponding image. For instance, the image can be used as a digital address to the lookup table, with the corresponding memory element describing the image. If n is as large as millions such large lookup tables cannot be built, and the question arises, whether we can replace such large lookup tables with smaller and manageable ones. In order to address this problem, we make use of the tools of Boolean algebra, which has a rich choice of established procedures and definitions.

We define a set of m objects (anything would do, like apples, oranges, rectangles). They form a 2^m element Boolean algebra (here and in the following we assume that also the position in space is characterizing these objects. For instance a rectangle in one certain position and an otherwise identical rectangle at a different position in space are considered as two different objects.
Consider now the images of these objects. In general, the images will not form a 2^m element Boolean algebra, because the objects may overlap, and the image of a partially occluded object is different from the image of the completely visible object. But this is not a

problem of image reconstruction in the very sense – rather it is due to the fact that images are two-dimensional, and the physical world has three space dimensions. In a two dimensional physical world overlap would not exist. One might refer to this as a problem of representation, but it is not a problem of message processing.
In a first approach we therefore choose objects which do not overlap, in order to form the set with m elements (as mentioned before, the position of an object is a property of the object). Then, both the original objects and their images form Boolean Algebras with 2^m elements and are isomorphic.
Furthermore, if we have associated to each image its correct description by means of a lookup table objects, images and their descriptions would form Boolean Algebras of 2^m elements.
If a_i is one of the objects, and if $b_i=D(a_i)$ is the description of the object, it is therefore true that the description of an image showing two objects, a_i and a_j, is identical to the list of the two descriptions for each single object: : $D(a_i \vee a_j) = D(a_i) \cup D(a_j)= b_i \cup b_j$.
This is how and why we can dramatically reduce the size of our universal image reconstruction lookup table: instead of describing all possible images of m objects with a lookup table of 2^m elements, we can describe them equally well with a lookup table of m elements.
Note, that this reduced lookup table is not an approximation. Rather it is analytical precise, like for instance the device in figure 3. In this sense, the fact that images form Boolean Algebras allows us to reconstruct a universal image reconstruction device, which describes images with the same analytical precision with which we can calculate $a+b = c$. We are now able to calculate the description of images. The process of "seeing" has become a calculus.

In order to build a real image processing system, we still need to consider two things:
Our argument does not only apply to sets of apples and oranges, but instead we can use as elements (Boolean atoms) any kind of object, like for instance line segments. The problem then arises of how to combine these segments to build larger objects, like straight lines or curves - and this should not be done by means of geometrical or phenomenological descriptions, but again using the methods of Boolean algebra. This is possible by forming strongly ordered sets. By defining adequate ordering relations, line segments can be ordered into objects.
The second problem is how to deal with occluded objects, which previously have been excluded from the discussion. We can include them by using ideals. An ideal of an image of a rectangle is the set of all images of the rectangle with one or several pixels omitted. Instead of using images of non occluded objects we can use their ideals.
(We note that the term "ideal" sounds similar to Plato's "ideas" not by accident: an information processing system which disposes of the ideal of a rectangle cannot be distinguished from an observer which knows Plato's idea of a rectangle – both of them have the same capacity of recognizing incomplete rectangles.)

Along these lines the spin off company Isomorph srl has developed a software package, which allows to recognize complex objects (like human beings) in non controlled environments performing also a three dimensional reconstruction of the scene. The reconstruction program does not make use of motion detection. The software is described in detail in [8], Isomorph srl refers to it as "linear computing". Several demonstration movies can be found on the web site of Isomorph [10]. Considering that less than two man years of development work have been invested in this software, it is performing in a very satisfying way indicating the practical use of the theoretical scenario presented here. A first industrial application has been highly successful [10].

Ideally, one would perform linear computing not on a Turing machine, and in particular not on a von Neumann architecture. Rather one would prefer to use a “linear computing machine”. Operating on a digital input of n bits, an ideal linear computing machine would have 2^n signal lines and 2^n memory elements. Of course, such machines do not exist nowadays. A very interesting development towards such a linear computing machine does exist: IBM has recently presented the CELL processor. Using a system of sub processors, the CELL machine is able to operate on 1024 signal lines concurrently, and each of the sub processors has rather direct access to a dedicated memory space, so that it can very efficiently perform calls to lookup tables. The Isomorph srl software has been installed and executed successfully on this new processor with excellent results: the identification of a human being takes about 0.2 seconds, including three dimensional reconstruction.

MORE MARGINALIZATION OF UMANS?

In the past science has been accused to have marginalized the human being, fIrst by displacing it from the centre of the world onto an arbitrary planet somewhere in the universe. Then by showing its parenthood to the ape, then by showing that to a large extent humans are controlled by their subconscious, thus reducing the human being to one form of animal. (>>the science of evolution is clear ... that humans are animals<< wikipedia). This is not just an academic discussion, rather many people suspect that this marginalization is at least in part responsible for the disastrous crimes which occurred in the 20th century. For this reason many people have a critical and distant relation to modern science.

Recently, this development has considerably accelerated and turned into an even heavier underestimate of humanity. It has been claimed that the human brain is nothing but a poorly performing computer, thus moving the human being closer to a machine, which morally speaking is even less than an animal.

If we consider what has been discussed in this paper from a purely mechanical point of view, it appears that our discussion may support and accelerate further this marginalization. If we can program “seeing”, “understanding”, “recognizing” into a computer, then this seems to support the view of the human being as a machine.

We note instead, that in this paper we have merely tried to create the mathematical foundations on which arguments like the ones of Szilard must be founded in order to arrive at a physics theory of information. This is just the beginning of a discussion on the physical theory of information.

As a next step, thermodynamics must be considered, as Szilard had attempted to do.

According to the second law of thermodynamics, only processes can occur, which increase the entropy of a closed system (this closed system can for instance be the Universe). Vice versa, processes which increase the entropy of the Universe will spontaneously occur, provided they are allowed by the law of physics.

If Szilard was right, if energy must be dissipated when importing information into an information processing system, then one would deduce, that the “diffusion” of information into the brain would be a spontaneous process. To the best of our knowledge, no experimental research exists as regards to this question, and therefore no experimental evidence. But it is interesting to note, that this question (whether the diffusion of information into the brain is dissipating energy and whether it is therefore spontaneous) should be accessible to the scientific method, it is a question of scientific nature. Further it is interesting to note, that a process of spontaneous diffusion of information into the brain would help to explain cultural evolution. It also would define a direction of evolution. As a

consequence one would conclude, that human beings are more highly developed than animals (this conclusion is totally incompatible with present evolution theory) because they represent more information than animals.
This would undo the seeming historical marginalization of the human. And it would do this, without contradicting Darwin, Freud or modern evolution biology. It would not be in contradiction, because up to now science was dealing only with the material world, for instance with organic live. And of course, with respect to its organic live, the human being is not different from the other animals.
And even without considering the hypothesis of a spontaneous process of diffusion of information from the outside world into the brain, another conclusion still can be found, again in contrast to the idea, that the human be marginal: if the vector scenario of information is true, then our brain cannot be anything else than an mirror of information found in the outside world. Cultural development would be a development towards a state of isomorphy between our brain and the world. Regardless of what the precise mechanisms are which lead to consciousness – we at least could say, whose consciousness our consciousness would be. If our brain is a mirror of the outside world, our consciousness necessarily would be the consciousness of the world. In this scenario, the Universe would exist also without us, but without us it would not be conscious of itself. In this scenario, we are the consciousness of the Universe. In this sense the old picture of the human as "centre of the world" would be re established and justified.

Much theoretical and experimental work still needs to be done. From a physics point of view this work is worth doing, since the physics of information gives us for the first time the chance of discussing thought itself in terms of physics – how and why we are doing physics becomes a subject of physics. From a more general point of view these studies should be supported, since the physics of information seems to have the potential to undo the seeming marginalization of the human being.

This paper has been made possible by the interest of Giorgio Bellettini. Many encouraging discussions with Gregory Chaitin have convinced us to insist on the project. The support of *Isomorph srl* is gratefully acknowledged

References

[1] C. E. Shannon: *A mathematical theory of communication.* Bell System Technical Journal, vol. 27, pp. 379–423 and 623–656, July and October, 1948.
[2] Ray Solomonoff, A Formal Theory of Inductive Inference, Part I", Information and Control, Part I: Vol 7, No. 1, pp. 1-22, March 1964
[3] C.S. Wallace, D.L. Dowe, *Minimum Message Length and Kolmogorov Complexity*, Computer Journal, Vol. 42, No. 4, 1999
[4] G.J. Chaitin, Algorithmic information thery, IBM Journal of Research and Development 21 (1977), pp. 350–359, 496
[5] L. Szilard, *über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen*, Zeitschrift für Physik **53**, 840-856.
[6] R.Landauer, *Irreversibility and heat generation in the computing process*, IBM Journal of Research and Development, vol. 5, pp. 183-191, 1961
[7] C.H.Bennett, *Logical reversibility of computation,* IBM Journal of Research and Development, vol. 17, pp 525-532 (1973);

C.H.Bennett, *The thermodynamics of computation – a review*, Int.J.Theor.Phys. 21, pp 905-940 (1982);
C.H.Bennett, *Notes on the history of reversible computing*, IBM Journal of Research and Development, vol. 32, pp 16-23 (1988)
[8] H. Grassmann, *On the mathematical structure of messages and message processing systems*, Isomorph Letters (A) 1 (2007). http://letters.isomorph.it
[9] Wu-Ki Tung, *Group Theory in Physics*, World Scientific Publishing Co., 1999
[10] www.isomorph.it