Computational Entropies

José Manuel Fernández
Department of Computer Science
University of Toronto
Toronto, Ontario M5S-1A4
Canada

Abstract

We present an example of how certain paradigms of Physics can be used in Computational Theory. More concretely, we show how the models of Thermodynamics can be generalized, and how this generalization can be used to define a new Theory of Information where computational restrictions matter. Our ultimate objective is to come up with a conceptually sound and usable definition of entropy in the computational context. We informally discuss how this can be done.

1 Introduction.

The main underlying principle of the traditional Physics of Computation has been the fact that all computation happens in the Physical Universe, and thus that the laws of information and computation depend on those of physics. In other words, that since the computing devices we built are a part of this world, a comprehensive Theory of Computation should not ignore the laws governing them and their physical limitations. This has probably been the central critique of the Theory of Computation by the physical field.

But beyond this principle, there is still much in Physics that is of interest in Computational Theory, for example, the mathematical language used in Physics, or the "physics-like" Mathematics. Roughly speaking, we can "import" physical quantities from Physics, that would allow richer descriptions of computational phenomena.

Entropy is a prime example of a successful use of this idea, having been defined first in Thermodynamics and then in Information Theory. Eventually, the relationship between both definitions was established (see Brillouin [1]), even though its formalization and complete understanding is still a topic of discussion, and will be, even at this conference.

However, in the context of modern Theory of Computation, the initial definition of entropy as given by Shannon [3] has its limitations. The main problem with Shannon entropy is that it measures quantities of information that are often incomputable if we have limited computational resources. In other words, the computational limitations of an observer often translate into a loss of information, that the Shannon entropy does not reflect. For example, the Shannon entropy of a message after encryption is zero, since the message can be reconstructed uniquely (by finding the decrypting key). However, if the encryption scheme is computationally secure, an observer with limited computational power cannot find out the key nor the initial message, and thus the virtual entropy of the encrypted message is maximal, (i.e. encrypted message looks random).

In order to resolve this apparent contradiction, one has to realize that we are dealing with *two* different measures of entropy, one for an infinitely powerful observer, and one for an observer with limited computational resources. This idea of an entropy relative to the computational power of the observer, or *computational entropy*, was first introduced by Yao [4]. If computational entropy could be properly defined, then we could properly formalize and use the concept of information in Theory of Computation, something that is done all the time, but only at the intuitive level. It is in order to do this, that we introduce the following concepts.

2 Generalized thermodynamics.

Beyond its original conception as a "science of heat", we can view Thermodynamics as a particular way of studying dynamical systems. While in Classical Mechanics, for example, we assume that all the relevant physical magnitudes can be measure and taken into account, here we establish a hierarchy of such

0-8186-3420-0/93 $03.00

magnitudes. Some, intrinsically describe the internal state of the system but cannot be measured or taken into account directly, while the others represent an external or *observed* representation of the system. In physical terms, this translates into the dichotomy between the microscopic and the macroscopic worlds, where the intrinsic magnitudes correspond to energy levels of individual particles, while the extrinsic ones correspond to temperature, pressure, entropy, etc.

From this point of view, we can think of a generalization of the thermodynamical way of looking at things. A thermodynamics (lower-case) consists of a system being studied, a well defined internal behavior that describes it, and laws that govern its interactions with the exterior. For example, a gas constituted by perfect particles, kinetically interacting with each other according to the perfect gases assumptions, and a set of macroscopical magnitudes like temperature, pressure, volume, etc. which describe its macroscopic behavior.

3 Infodynamics: a thermodynamics of information.

In the case of Information Theory, the system being observed is what Shannon [3] called the *information source*. The information source is meant to be an abstraction of phenomena which produce information in the form of text (e.g. human languages). It is a "black box" which on request will produce symbols (one at a time) from a fixed alphabet, according to a certain probabilistic behavior. In the simplest case of information source, the *memoryless* information source, the probability of each of the individual symbols being produced will be a constant number. In the general case, the source is considered to behave as a generic Markov Chain, where these probabilities can depend on the symbols previously produced. The internal behavior of the source is determined by this model, while the observed or external behavior corresponds to the sequence of letters that have been output. For such a thermodynamics of information, the name *infodynamics* has been coined [2].

For ideal gases, the internal state of the system is determined when we know the individual positions and velocities (and thus corresponding energy levels) of all particles in the system. In the case of an information source, it is not so clear what an "internal state" corresponds to. However, if we knew in advance the sequence that it was going to output, we could later reproduce the same exact behavior of the information source to a third party, say, which could then not distinguish us from the initial source. That is why we adopt a "canonical representation" of internal states of an information source. Instead of thinking of an internal configuration of the source as being a sequence of non-deterministic choices of transitions between states of a Markov chain, we represent an internal state of a source as the (possibly infinite) sequence that it would output, if it was asked to do so forever.

This simplification allows a better mathematical treatment of entropy, and at the same time allows a certain independence in the choice of model; we do not have to assume then that our source behaves as a Markov chain of any particular kind, but simply that under different "microscopic" conditions, it will generate different (infinite) sequences of symbols. However, one of the disadvantages of this approach is that we may no longer assume that all possible (infinite) sequences are equally likely to be generated, i.e. that all microstates are equally probable.

In conclusion, while a gas with certain absolute parameters (number particles, possible energy levels, boundaries) can take different internal configurations, which correspond to particular assignments of particles to energy levels and positions, an information source also defined by certain fixed parameters (model, symbol probabilities, etc.) can adopt different configurations which correspond to a concrete (virtually infinite) sequence of symbols generated by it. Notice here how this is more of a temporal configuration, in contrast to a gas configuration which is an instantaneous, "spatial" configuration.

In Infodynamics the laws that regulate interaction with the exterior are easily enunciated. Action from the exterior is limited to requesting symbols, (and we have seen that this can be "factored out" by considering the virtual infinite sequence that will be output), and all that percolates to the exterior are the symbols output, and nothing else.

Alternatively, we can also use the notion of observer to describe Infodynamics (or any other thermodynamics), rather than using generic laws of interaction. In general, a thermodynamics can also be defined by the system being studied and the observer studying it, or more precisely the laws governing observations made by him. In the example of gases, the observer can only interact with the gas at the macroscopic level, i.e. extracting physical work or extracting information through measurement of macroscopic magnitudes (temperature, etc.). In the case of Infodynamics, an observer is someone who makes symbol requests to the source, and obtains all information about the system

exclusively through the symbols output (probabilities of symbols, previously output symbols, etc.).

This definition is more useful for our purposes. First it allows a unified characterization of entropy, and then it allows us to easily define a "good" thermodynamics for the computational context.

4 Entropy.

Entropy is the quantity of information on the system *not* available to the observer after observation; the amount of uncertainty that the observer has on the microscopic configuration of the system. It is thus an observer-related magnitude, or using traditional terminology a *macroscopic* magnitude. Let's be more specific.

We call *total information* of the system the amount of information needed to specify a particular state of the system, when nothing is known about it, i.e. no macroscopic observations have been made[1]. In the simpler cases where the system has finite states all equally likely, this quantity is simply the log base-2 of the number of states (measured in bits). This quantity is independent of the observer.

Still, the observer will be able to obtain some information about the source from his observations. He can measure temperature, pressures, can compute relative frequencies of the symbols output, etc. The maximum amount of information (measured in bits) that an observer could obtain, through any means allowed by the laws of observation of that thermodynamics is called the *partial information* of the system. This *is* an observer-dependent magnitude.

Thus, entropy is the difference between the total and the partial information.

5 Computational Infodynamics and Computational Entropy.

We can now understand why a same system can have different measures of entropy, as in the example of an encoded message in Section 1. These two entropy measures correspond to two different thermodynamics, which in turn are different because they correspond to two different observers: one which is infinitely powerful, and for which trying all possible encrypting keys is a "legal" means of observation; and the other one who must obey the additional rule that his observation cannot involve unreasonable computations.

The whole object of this discussion was to come with a proper framework for defining computational entropy. We can now characterize computational entropy as the entropy measure of a particular thermodynamics. This thermodynamics is that of information sources (the system being observed) where the observer is a probabilistic Turing Machine which runs in polynomial time[2]. There are many reasons to justify this choice of observer, but the main one is that it is believed in Complexity Theory that this model reflects the essence of what is computationally tractable. This thermodynamics we call *Computational Infodynamics*, in contrast with the traditional Infodynamics in which the observer has no computational restrictions.

Even though from the conceptual point of view, this characterization of computational entropy might be insightful, it does not provide us with a useful, complete definition. Part of the problem is that in this thermodynamics, it is hard to establish exactly what is the partial information that an observer can extract. It is defined as a *maximal* quantity, and in order to come up with a value for it we need to show that no observer can obtain more information, which is not easy given the generality of the Turing Machine model.

6 Axioms of Computational Entropy.

We try to bridge this gap between a conceptual and a useful definition by looking at the properties that computational entropy must have. Our hope was that we could "extract" enough such properties from our conceptual understanding of computational entropy, such that we could use these properties to define it axiomatically; which is why we called them "axioms." This approach has not been as successful as we had hoped, in that the properties found do not define computational entropy uniquely. However, these "axioms" do provide more insight about computational entropy. Informally presented, they are:

Axiom 1 *Let S be an information source. Let $H(S)$ be its entropy under the classical Infodynamics (aka, the Shannon entropy). Let $H_c(S)$ be the computational*

[1] this quantity is sometimes the *a priori* entropy.

[2] i.e. A Turing Machine that can flip coins, and whose running time does not increase asymptotically faster that a given polynomial whose free variable is the total information of the system.

entropy of the same source. Then,

$$H(S) \leq H_c(S)$$

This can be explained as follows. In computational Infodynamics, the observer has additional restrictions, that the observer in classical Infodynamics does not have. Hence, the classical observer can obtain more information, and its associated entropy can be no greater than that of its computational counterpart.

Axiom 2 *Let S and S' be two* computationally indistinguishible *information sources. This is, no observer in the computational Infodynamics can tell them apart by looking at their output sequences. Then we say the following:*

$$H_c(S) = H_c(S')$$

This axiom is the computational parallel of the fact that in classical Information Theory, two sources which are statistically indistinguishible (i.e. have the same symbol probabilities), have the same Shannon entropy. Note however, that even if two sources are computationally indistinguishible, they could have different Shannon entropies.

Using both axioms together we can provide a lower bound on the computational entropy of any source S, by finding another source S' computationally indistinguishible to it, whose Shannon entropy we know (or can compute, since Shannon entropy is well defined). Thus we have:

$$H(S') \leq H_c(S') = H_c(S).$$

Unfortunately, this is all we can do. These two axioms alone cannot provide us with an upper bound that would define computational entropy uniquely.

One way to achieve such an upper bound is as follows. Consider the Shannon entropies of all sources S' that are computationally indistinguishible to S. We say that the computational entropy of S is no higher than the maximum Shannon entropy of such sources. In other words, we make the computational entropy of S as small as possible, without violating axioms 1 or 2. This would be equivalent to introducing a "third" axiom:

$$H_c(S) \leq \sup_{S' \cong S} H(S')$$

Unfortunately, this would be quite arbitrary, and we do not know how to justify it conceptually. We do not even know if a measure of computational entropy so defined would sound, and obeying certain other desirable properties (e.g. an equivalent of Shannon's first theorem). Work in that direction is still necessary, as it remains an open question what a good "third" axiom would be.

7 Statistical Thermodynamics and Statistical Entropies.

Here is another approach to define computational entropy, more along line of how Shannon entropy is viewed in classical Information Theory.

Let's set aside for a moment the idea of observer, and come back to the idea that what characterizes a thermodynamics is the nature of its microscopic reality vs. its macroscopic reality. In purely mathematicals terms, a thermodynamics is a triplet consisting of a set of microscopic states, a set of macroscopic states, and a (onto) function of observation, mapping the ones onto the others. While the microscopic states represent the internal or "real" behavior of the system, the macroscopic states represent the behavior from the observers point of view. Obtaining total information on the system is equivalent to determining the microstate it is in. However, the observer can at most determine which macrostate it is in. Equivalently, all microstates that are indistinguishible to the observer will map unto the same macrostate. More formally, the macrostates form a partition of the microstates, whose equivalence relation is indistinguishibility by the observer.

Classical Thermodynamics and Information Theory share a common assumption. We are assuming that given a certain distribution of the microstates, normally the uniform one (i.e. that they are all equally likely), there exists a proper probability distribution of the macrostates induced by the observation function. This is quite easy to see in the finite case; the probability of a particular macrostate is proportional to the number of microstates that map unto it. This assumption allows a nice natural definition of entropy, i.e. the one introduced by Shannon [3], which is the one everyone knows about:

$$H(S) = \sum_m \Pr(m) \log \frac{1}{\Pr(m)}$$

where S is the system/source being studied, and m ranges over all possible macrostates.

Let us try to use the same approach in defining computational entropy. Here the macrostates correspond to subsets of (infinite) sequences generated by one information source, which are all computationally indistinguishible with each other. In this case, it is not clear what a probability distribution on these macrostates would be, if it exists at all.

It was initially believed that taking such an approach was not possible, as we did not think probability distributions could be defined for computational

Infodynamics; in a very vague sense it seemed to us that information sources behaved "non-ergodically" under computational observation. This is why we initially tried the "axiomatic" approach detailed above. However, we have found one approach that could possibly allow us to define such distributions, nonetheless.

Let us imagine that we can come up with a basis for all sequences, such that all possible sequences would be computationally indistinguishible to one (and only one) of the sequences of the basis; that way each subset of computationally indistinguishible sequence would have a representative. Then the probability distribution of a given macrostate would be given by the probability that its representative and a "random" sequence could be confused by the observer. If such probabilities can be properly defined on the domain of probabilistic Turing Machine running in polynomial time, then we could use the same statistical definition of entropy used in classical Infodynamics. To this day, this remains a possibility only, and has not been formally described yet.

8 Conclusions

We have presented here a new framework in which the conceptual definition of computational entropy is much clearer and sound that previous attempts ([4]). The vocabulary and way of thinking is of value by itself, as it shows us how it is possible to give better descriptions of concepts that appear in Computational Theory, thus broadening our understanding of them.

We have also presented two different attempts at using this new way of thinking in a more formal setting, to give concrete, usable definitions. Unfortunately, to this day, these attempts have not been completely successful. However, we think that in the future additional work on this topic will prove fruitful.

References

[1] Brillouin, L. 'Science and Information Theory', Academic Press, New York, 1962.

[2] Fernández, J.M., "Entropy and Computation", *Bachelor's thesis, Mass. Inst. of Tech., Dept. of Comp. Sc.*, 1991.

[3] Shannon, C., "A Mathematical Theory of Communication", *Bell Systems Technical Journal*, 27, 1948, pp. 379–423 and pp. 623-656.

[4] Yao, A.C., "Theory and Applications of Trapdoor Functions (extended abstract)", *Proc* 23^{rd} *IEEE Symposium on the Foundations of Computer Science*, 1982, pp. 80-91.