

Article

Physical Intelligence and Thermodynamic Computing

Robert L. Fry

Applied Physics Laboratory, Johns Hopkins University, Laurel, MD 20723, USA; robert.fry@jhuapl.edu

Academic Editor: Dawn E. Holmes

Received: 22 January 2017; Accepted: 6 March 2017; Published: 9 March 2017

Abstract: This paper proposes that intelligent processes can be completely explained by thermodynamic principles. They can equally be described by information-theoretic principles that, from the standpoint of the required optimizations, are functionally equivalent. The underlying theory arises from two axioms regarding distinguishability and causality. Their consequence is a theory of computation that applies to the only two kinds of physical processes possible—those that reconstruct the past and those that control the future. Dissipative physical processes fall into the first class, whereas intelligent ones comprise the second. The first kind of process is exothermic and the latter is endothermic. Similarly, the first process dumps entropy and energy to its environment, whereas the second reduces entropy while requiring energy to operate. It is shown that high intelligence efficiency and high energy efficiency are synonymous. The theory suggests the usefulness of developing a new computing paradigm called *Thermodynamic Computing* to engineer intelligent processes. The described engineering formalism for the design of thermodynamic computers is a hybrid combination of information theory and thermodynamics. Elements of the engineering formalism are introduced in the reverse-engineer of a cortical neuron. The cortical neuron provides perhaps the simplest and most insightful example of a thermodynamic computer possible. It can be seen as a basic building block for constructing more intelligent thermodynamic circuits.

Keywords: Carnot cycle; causality; distinguishability; entropy; intelligent processes; questions

1. Introduction

Work on a fundamental theory of computation that may underlie intelligent behavior has been pursued over many years. This paper introduces that theory and highlights aspects of an emerging engineering formalism.

The term *computation* enjoys a wide range of meaning. Moreover, what can perhaps be agreed upon is that computation is the conversion of an input into an output. This paper provides a precise philosophical and mathematical formulation of what comprises computation as carried out by physical systems and processes. This in turn leads to an explanation of intelligent processes that is entirely physics-based.

This theory of computation is based on two fundamental axioms. These pertain to distinguishability and causality. These comprise the two most basic of all physical phenomena. The theoretical implication of distinguishability is a generalized form of Boolean algebra that includes conventional logic and by duality, an algebra of logical questions. Logical questions have significant engineering utility in that they provide a powerful way of symbolically capturing and manipulating intelligent process concepts. The consequence of causality is that there are only two kinds of computational dynamics possible—one that includes intelligent processes and another that includes dissipative physical processes for which communication systems serve as an example.

Learning and adaptation within intelligent process can be understood from two mathematically equivalent perspectives. First, it can be seen as the system optimizing its computational efficiency. Also, through a simple transformation of its objective function, it can be seen as maximizing its energy

or Carnot efficiency. From a computational perspective, the system is solving a min-max optimization problem that simultaneously balances the rate at which it acquires information to the rate it makes decisions. From a thermodynamic perspective, it is attempting to solve a min-max Helmholtz free energy problem that optimizes its energy efficiency.

Current research focuses on the development of a formal engineering framework for what is called a *thermodynamic computer* or TC. Theory provides insights into how to engineer TCs. A principal pathfinder application has been the reverse-engineering of the cortical neuron [1].

Section 2 develops the underlying computational theory. It is the most difficult discussion both conceptually and philosophically. It requires the reader to think quite differently about what comprises intelligence and autonomy. Section 3 focuses on mathematics and develops the statistical dynamics and physics of intelligent processes and TCs. Much of this is derived from developments in information theory. It shows how basic thermodynamic principles can underlie intelligent processes. Section 4 provides a TC-based explanation of how cortical neurons work and adapt. Section 5 concludes by outlining critical topics not covered in this paper.

2. Physical Intelligence and Computation

This paper proposes a physical basis for intelligent processes, provides examples, and suggests that a formal engineering framework for engineering intelligent systems is possible. Theory requires consideration of a new form of computation that can provide a common basis for understanding intelligent and dissipative physical processes as a whole. It is based on quantifying causality and what it means to subjectively distinguish one's environment from the standpoint of computation. This section contains the following subsections:

- Distinguishability
- Causality
- Type I Computation: Communication systems
- Type II Computation: Intelligent systems

2.1. Distinguishability

At a minimum, computation requires two things: logical tokens of computations (i.e., things that are operated on and that change) and rules that dictate the dynamics of how tokens evolve. The proposed axioms meet these requirements by defining the tokens of computation in the first and its possible dynamics by the second.

Any intelligent process is attempting to solve a problem. Whatever the problem, it corresponds to what is called its computational issue. An intelligent process wants to solve its problem by formally resolving its issue through its processing.

A practical example is that of finding one's car keys for the drive to work in the morning. The issue is: "Where are my keys?" They are lost! Resolving this issue requires searching every room in the house to hopefully resolve the problem by finding them. A smart search strategy would be to first search the most likely rooms where the keys could be (e.g., entryway, bedroom, kitchen). This will, on average, reduce search time and minimize the energy required to solve this problem.

As will be shown, one can quantify intelligence as an exact number specifiable in units of bits per decision by way of analogy to the bandwidth of digital communication systems [2]. It is the rate given in units of bits per second of the capacity of a system to resolve its computational issue. The faster a system can solve its problem, the higher its intelligence.

Consider the following two axioms:

1. A1: The ability of a system to distinguish its environment is the most basic operation any system can perform.
2. A2: Computational dynamics must abide by causality.

A1 is discussed first. It captures how a system can distinguish its environment and what subjectively comprises information and control.

Consider a system that can observe *X* and make decisions *Y*. Uppercase italicized letters denote logical questions. By asking *X*, the system can acquire information from and about its environment. By making decisions *Y*, it can control it.

Consider a simple single-cell organism, a protozoan. Some protozoa have the ability to sense light and to “see”. They can also propel themselves in water using their cilia [3]. Consider such a species living in a pond and that it likes to eat algae and other material found near the pond surface. Water near the surface is preferentially illuminated by the sun and tends to provide a richer source of nutrients that the protozoa like to consume.

Imagine protozoa in the pond being perturbed by internal currents that randomly reorient them over time. Moreover, there will be times when a protozoan is oriented such that its forward optical sensor will point in the direction of the brighter pond surface. It may or may not sense light as dictated by its detection threshold and its depth in the pond. The organism’s ability to sense light corresponds to the subjective question *X*. Possible answers to *X* must correspond to a binary internal state of the organism placed into one state or another by its environment.

Upon a positive detection, the protozoan may or may not activate its cilia to propel itself forward and toward the surface of the pond where it can successfully feed. Call this decision *Y*. It again is an internal state set of the organism with it representing one of two possible binary decisions. However, in this case, it is the organism, not the environment, that decides this state.

The only energy-efficient survival strategy for the protozoan is for it to move toward light when it sees it. However, there are three other possible strategies that can be delineated: (1) never activate its cilia when light is detected; (2) activate its cilia when it does not sense light; or (3) never activate its cilia under any circumstances. None of these support survival. Therefore, in some regards, one can say that this simple creature behaves intelligently in how it converts information into action. It could not flourish otherwise.

Now associate the physical protozoan system with a computational issue denoted *A* to survive. To sense light, *A* asks the question $X \equiv \text{“Do I sense light?”}$ The induced response within *A* is that this question is either answered “yes” or “nothing” (i.e., no response from its threshold detector). To move, *A* can be thought of answering a question $Y \equiv \text{“Should I activate my cilia?”}$ by either doing so or not through the activation of an internal state.

In general, a subjective inquiry *X* allows a system *A* to actively acquire information from its environment. Answers are determined by the specific question posed and by the environment. However, the ability to pose *X* is defined totally within the system *A* itself. Most fundamentally, it is how *A* distinguishes its environment from an information perspective. Similarly, answering *Y* allows *A* to make subjective decisions as to how it wants to behave in its environment. It is the way *A* subjectively distinguishes the possible ways it can control its environment through its subjective decisions. Note that philosophically a question represents a way that a system establishes boundaries of its subjective reality. The world is as we apportion our subjective experience into discrete experiential quanta. The philosopher Clarence Irving Lewis [4] first used the term “qualia” to capture this idea.

The notion of distinguishability underlies the subjective acquisition of information by an intelligent system and its use to make decisions. If one could formalize the notion of questions in this manner, it would allow for a powerful computing paradigm by providing a formal way of considering what it means to ask and answer questions within the subjective frame of a system.

Cox [5], as inspired by Cohen [6], showed that Boolean algebra as currently known is incomplete. In fact, exactly half is missing. Cox formulated a joint Boolean algebra of assertions and questions that captures the elementary dynamics of the physical exchange of information and control of a system with its environment.

Conventional Boolean algebra is an algebra of propositions. “This ball is red!” Conversely, Cox [5] proposes two mutually defining quantities he calls assertions and questions. “An assertion

is defined by the questions it answers. A question is defined by the assertions that answer it.” These definitions result in two complete and complementary Boolean algebras, each with its respective logical measures of probability and entropy. Assertions, questions, probability, and entropy all have physical manifestations and are intrinsic to all dynamical system processes including intelligent ones.

That is, logical propositions in fact assert answers or states of nature like “The ball is red!” Every logical expression in conventional logic has a dual question, e.g., “Is the ball red?” Action potentials can be thought of as computational tokens in brains that answer questions posed by the dendrites of neurons or answered by their output axons. Likewise, photons are assertions that answer the question posed by an optical detector.

As a trivial example, consider a card-guessing game where one player desires to know the card held by another. Lower-case italicized letters denote logical assertions such as a and b . The first player resolves this issue by asking the second player questions like $S \equiv$ “What is the suit of the card?” or $C \equiv$ “What is the color of the card?” These questions are defined by the assertions that answer them, where $S \equiv \{h, c, d, s\}$ and $C \equiv \{r, b\}$ to make explicit the fact that questions are defined by the assertions that answer them. Of course $h \equiv$ “Heart!”, $c \equiv$ “Club!”, $d \equiv$ “Diamond!”, and $s \equiv$ “Spade!”, whereas $r \equiv$ “Red!” and $b \equiv$ “Black!”.

The common question is denoted by the symbol “ \vee ” with $S \vee C = C$ an example of usage. This is the disjunction of two questions and asks what both question asks. The information common to suite S and color C is the color of the card C . Intelligent systems are interested in $X \vee Y$, the information X common to the decisions Y to be made. In the case of the protozoan system, $X \vee Y$ is all the organism is interested in because this drives its survival. Disjunction is a powerful and simple construct for quantifying the concept of actionable information.

The joint question $S \wedge C = S$ provides the information given by asking both questions S and C . This is the conjunction of two questions and in this case is the suit of the card question $S = S \wedge C$.

Regarding disjunction [5], the question $S \vee C \equiv \{h, c, d, s, r, b\}$. Moreover, any assertion that implies an assertion that defines a question also defines that question. In this case, $h \rightarrow r$, $d \rightarrow r$, $c \rightarrow b$, and $s \rightarrow b$, and so $S \vee C \equiv \{r, b\} \equiv C$. Regarding conjunction [5], $S \wedge C \equiv \{h \wedge r, h \wedge b, c \wedge r, c \wedge b, d \wedge r, d \wedge b, s \wedge r, s \wedge b\}$. Assertions like $h \wedge b$ and so on where suites and colors are absurd and can be eliminated leaving $S \wedge C \equiv \{h \wedge r, c \wedge b, d \wedge r, s \wedge b\}$. But for instance $h \wedge r \rightarrow h$ and so $S \wedge C \equiv \{h, c, d, s\} \equiv S$.

In the case of the card-guessing game, a special logical relationship exists between C and S called strict logical implication [7], whereby “ S implies C ” or $S \rightarrow C$. Either condition $S \vee C = C$ or $S \wedge C = S$ is necessary and sufficient for strict implication to hold. The implication of one question by another means that if S is answered, then so is C . In the case of the protozoan system, it desires the condition $Y \rightarrow X$ so that it properly activates its cilia when it sees light.

Assertions can also strictly imply one another. An assertion a implying another is denoted by $a \rightarrow b$. This means that if b answers a question, then so does a . The assertion a may contain information extraneous to answering the reference question. For example, $s \equiv$ “He is my son!” implies $b \equiv$ “He is a boy!” regarding the gender question $G \equiv$ “Is your child a girl or boy?” Upon hearing “He is my son!” the person posing question G will strip away the extraneous information about “son” and only hear “boy” given the inquiry $G \equiv \{b, g\}$.

In probability theory, the probability of c given a is denoted by $p(c|a)$. Formally, probability is the measure or degree [8] to which one assertion a implies another c or $a \rightarrow c$. The introduction of notation $p(c|a) = (a \rightarrow c)$ denotes this. By way of symmetry, one can ask if there is a similar measure for the degree of implication of one question by another, and indeed there is. Cox [5] introduces the notation $b(B|A)$ as the bearing of one question B on answering another A , which is represented by the notation $b(B|A) = (B \rightarrow A)$. Formally, bearing b is just entropy [5]. The protozoan, for instance, wants to maximize $(X \vee Y \rightarrow A) = b(X \vee Y|A)$, where A is formally its computational issue to survive. In the card-guessing problem, $b(S \vee C|A) = b(C|A)$ and $b(S \wedge C|A) = b(S|A)$. The first provides one bit of information, and the second provides two bits.

The complement of a question $\sim B$ asks what B does not ask relative to resolving an issue A . Therefore, $\sim B \wedge B$ asks everything. Similarly for assertions, $\sim b$ asserts what b does not, and this is the active information perspective assumed. In this case, $\sim b \vee b$ tells everything and is always asserted.

It is easy to map back and forth between logical expressions involving bearing and equivalent information-theoretic expressions. This mapping is summarized in Table 1.

Table 1. Relationships between bearing and information-theoretic measures.

Bearing	Information Theory	Name
$b(X A)$	$H(X)$	Entropy
$b(X \vee Y A)$	$I(X; Y)$	Mutual Information
$b(X \wedge Y A)$	$H(X, Y)$	Joint Entropy
$b(X \vee \sim Y A)$	$H(X Y)$	Conditional entropy

The remainder of this paper emphasizes the use of the notations $I(X; Y)$, $H(X)$, and $H(Y)$ instead of $b(X \vee Y|A)$, $b(X|A)$, and $b(Y|A)$, respectively. However, keep in mind that any particular intelligent process has a reference frame A to which the corresponding information-theoretic expressions refer. For example, $H(X)$ is sometimes called “self-information”. Conversely, $b(X|A)$ measures the amount of information to be expected by posing the question X to resolve an issue A . In the card-guessing game, C and S are posed relative to the broader issue $A \equiv \{c_1, c_2, \dots, c_{52}\}$ of which card c_i in a deck of 52 cards is being held by the other player.

Venn diagrams [9] provide a convenient way of graphically capturing strict logical implication of assertions and questions. If $b \rightarrow c$, then a smaller circle labeled a will reside somewhere inside a larger circle labeled c such as depicted in Figure 1d. If $B \rightarrow C$, then a larger circle B contains a smaller circle labeled C again as depicted in Figure 1d right bottom. This depicts that the information request of B is greater than C and includes that of C . The latter diagrams are formally called I -diagrams [10], whereas those for assertions are conventional Venn diagrams.

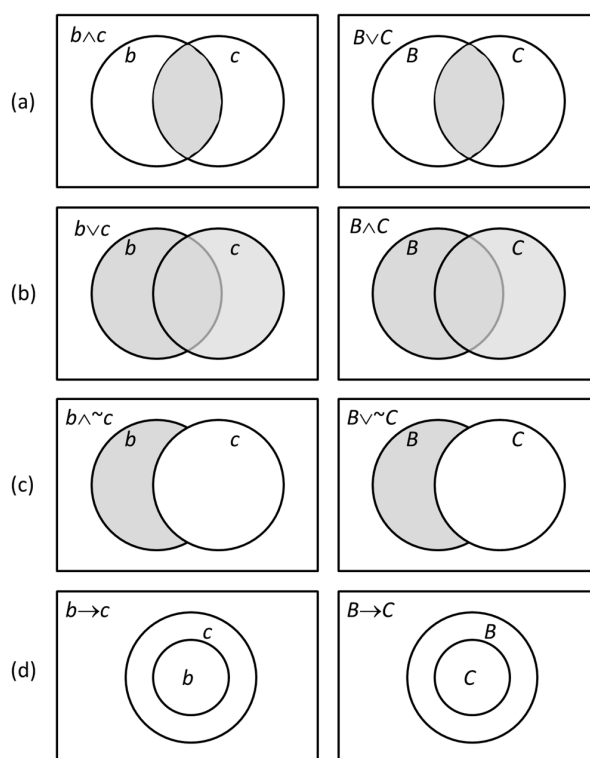


Figure 1. Comparison of dual Venn diagrams (left) and I -diagrams (right).

Figure 1 illustrates dual Venn and I -diagrams for several cases of interest. The transformation between each column is accomplished using the reflection operator [11] where, for instance, $\overline{a} = A$, $\overline{A} = a$, and $\overline{a \vee \sim b} = A \wedge \sim B$. The conversion rules [5] are that assertions and questions are interchanged, as are conjunction “ \wedge ” and disjunction “ \vee ”. The complementation operator is preserved.

The I -diagram in Figure 1d is of particular interest. Suppose that an intelligent process asks $X = C$ and decides $Y = B$. As will be shown, the condition $Y \rightarrow X$ means that “control authority” of the system exceeds its informational uncertainty. This implies $b(X \vee Y | A) = b(X | A)$, meaning that the available information is adequate to provide for perfect control.

As an example, consider the shell game played at least as far back as ancient Greece [12]. In this game, one player using sleight of hand quickly moves three nut shells around one another. One of these shells contains a pea. The objective of the second player is to guess which shell contains the pea. In this game, suppose that the second player has the information $X \equiv \{x_1, x_2, x_3\}$. If this player has good visual acuity, he may be able to follow and track the moving shells. In general, this is not the case. The second player then makes the decision $Y \equiv \{y_1, y_2, y_3\}$. Because of inadequate information, the second player cannot make perfect decisions. If, however, the second player had the option of making additional choices following wrong ones, he would be guaranteed the ability to win the game given $Y \rightarrow X$. This increase in control authority is required to overcome informational uncertainty.

Consider a physical example of a question and assertion consisting of a photon detector with quantum efficiency η . If a photon hits the detector, it is detected with probability η . The detector corresponds to a question $D \equiv$ “Is a photon present?” Photons $d = \overline{D}$ correspond to assertions. If one can instead physically build a “photon absent” detector, it would be denoted by $\neg D$ and have the property that it would only respond if no photon is detected.

Photon detectors, photons, dendrites, and axons and action potentials are examples of binary questions. They can be graphically described in a diagram called a dyadic, as shown in Figure 2 for $Q \equiv \{a, \neg a\}$. The complementation operator allows the transformation of assertion states, whereas the reflection operator transforms between corresponding assertions and questions.

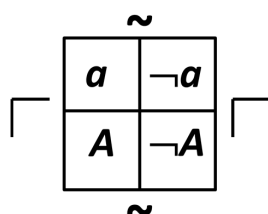


Figure 2. Dyadic diagram used for characterizing a binary question.

Cohen [6] described the importance of having a formal means of logically manipulating questions by way of analogy with conventional logic. Clearly, he had some anticipation of Cox’s work [5] as reflected by his statement in Reference [6]: “In order to be able to ask the question ‘What is a question?’, one must already know the answer”.

2.2. Causality

So, what does it mean that computation must abide by causality? Causality has many interpretations. The one assumed here is both physical and computational. Answering the question “What is causality?” seems equally elusive to answering “What does it mean to distinguish?”.

Fortunately, an answer is found in an obscure, but profound, statement that Shannon made in what is perhaps his most famous and important paper [13] in 1959:

“You can know the past, but not control it. You can control the future, but have no knowledge of it.”

Shannon spoke of the importance of this observation and that he would write more about it. However, for unknown reasons, he never did. Information theorists have unsuccessfully sought to understand exactly what Shannon meant by those statements [14].

One is compelled to believe that Shannon saw important implications in what he said. Exactly what he thought these were will remain forever unknown. However, one is certainly granted the freedom to speculate.

Assume that the implication of Shannon's causality is that two kinds of computation are possible. Processes can exist that reconstruct the past, or other processes can exist that control the future. As will be shown, communication systems [2] provide good examples of the first kind of system. Conversely, it will be shown that intelligent systems attempt to control the future. It is interesting that the latter have a much simpler explanation than the first, with both being complementary and dual to one another in ostensibly every way.

Before beginning, both kinds of systems, whether they are concerned with the past or the future, perform their computations in the present. Not doing so would violate causality. Also, both kinds of processes possess physical origins to which system entropy or uncertainty is referenced. Otherwise, as described by Gibbs [15], there is no thermodynamic need for an absolute origin. For instance, the Carnot cycle of the neuron will be seen to shift its position in location 1 bit to the left or right on an entropy plane as the number of its dendritic inputs changes by 1.

For a communication system, the origin physically and functionally corresponds to the receiver. The origin of an intelligent system corresponds physically to the location where acquired information is converted into decisions. Communications systems are discussed first.

2.3. Communication Systems

Consider a process that reconstructs the past. A communication system does this because its only purpose is to reconstruct information transmitted through a channel. It tries to reconstruct at the present a symbol transmitted to it in the past by a source with a time delay specific to the channel being used. The system must be synchronized such that the receiver will know when the source transmits a symbol and when it should expect to receive it. The symbol itself is not directly received; instead, a modulated signal is inserted into the channel by the source. This signal, while propagating within the channel, is subject to information loss relative to the receiver caused by possible attenuation, channel distortion effects, and noise.

Suppose that the communication system has an input alphabet given by x_1, x_2, \dots, x_n . The receiver asks the question $X \equiv$ "What symbol was transmitted by the source?" Formally, $X \equiv \{x_1, x_2, \dots, x_n\}$. As the receiver awaits reception, it possesses uncertainty $H(X)$ regarding what was sent. The entropy $H(X)$ is dictated by the alphabet of the communication and the relative probabilities the receiver assigns to each symbol it deems possible to have been sent. The key item of note in this perspective is that the uncertainty $H(X)$ is defined relative to the receiver because the source knows perfectly well what it is transmitting, whereas the receiver does not.

The source, having selected a symbol to send, will modulate and expend energy to transmit it through the communication channel to a receiver. While propagating through the channel, the energy of the signal will be passively dissipated to the environment. At the same time, the signal will be affected by channel noise and distortion effects. These effects combine to reduce the ability of the receiver to perfectly reconstruct codes sent to it in the past.

Depending on the physical proximity of the source and the receiver as measured through the channel, the receiver will measure or otherwise acquire the signal observed at its end of the channel. The receiver necessarily must include an algorithm that will demodulate this signal and then compute a single number that guides its selection of a symbol from its alphabet of hypothetical symbols it thinks could be sent to it by the source in the past. This serves to reduce its entropy $H(X)$.

This number must causally exist prior to the receiver's choice. Hence, this number must be stored in some type of memory and be causally available when this decision is made. Once this specification is

made by the receiver, the utility of the acquired and stored information is consumed. Once exhausted, the entropy of the receiver is restored to its original value $H(X)$. This restoration occurs at the receiver.

Once this occurs, two other actions take place. First, the receiver formulates a new inquiry as to its expectation of the next received code. This formulation can occur instantaneously once a current received code is selected and serves to restore $H(X)$. Secondly, the memory of the receiver must be ready to record the next measurement to be used in selection of the successive symbol. All of these computational elements are captured by a Carnot cycle depicted on a temperature-entropy plane, as shown in Figure 3.

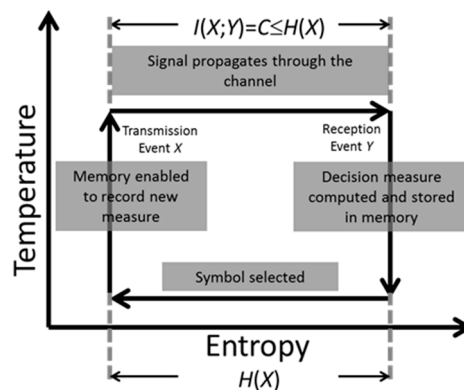


Figure 3. Information transmission process depicted as a Carnot cycle on the temperature-entropy plane.

As shown in Figure 1, if the channel capacity C equals the source entropy $H(X)$, the receiver can perfectly reconstruct the past. Otherwise, there will be transmission errors.

There are many implications of adopting a process-oriented perspective of information theory, although not all are covered in this paper. However, assuming the perspective that communication processes can be treated thermodynamically has engineering utility as to what should be explored further.

2.4. Intelligent Systems

Intelligent processes attempt to control their future. They operate in a complementary manner to the communication process previously described. However, their operation is in many ways more intuitive.

Communication processes are exothermic in that energy and information are lost during signal propagation through the channel. The area of the Carnot cycle in Figure 3 given by $\Delta T \Delta H$ equals the energy dissipated [16]. To overcome this, symbols have to be successively reconstructed through re-amplification and then retransmission. This requires energy. As an example, one need only think of undersea fiber-optic channels with repeaters that periodically re-amplify and retransmit codes. This process requires an energy source to ensure continued error-free transmission of signals across oceans.

Conversely, intelligent processes require energy to work and compute. They are endothermic. Hence, they must operate in a thermodynamically inverse manner to communication processes and therefore can be thought to operate as depicted in Figure 4.

Figure 4 also describes the computational phases comprising a single process cycle within an intelligent process. First, the system poses a question X to its environment. Information is acquired at a constant “high” temperature. This means that the information in the acquisition channel is subject to measurement noise just like that in a communication channel.

The information in X common to the decision Y is logically given by $X \vee Y$. This is formally the actionable information the system possesses relevant to making its decision Y . It takes the form of a single numeric quantity derived from measured data. This mirrors the subjective measure computed

by the receiver of a communication system. Once computed, and also as in a communication system, it is stored in memory and the system transitions to a lower temperature. Physically, this means that the actionable information is a fixed numeric quantity stored within the system where it is no longer subject to noise.

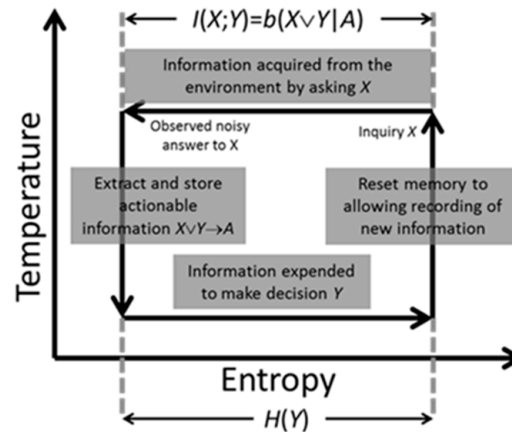


Figure 4. Carnot process associated with an intelligence process as depicted on the temperature-entropy plane.

The process of storing information is analogous to a physical system undergoing a phase transition because of a reduction in temperature. That, in turn, implies a reduction in the number of thermodynamic microstates that the system can reach. This, for instance, occurs when a liquid transitions into a solid.

Once actionable information is available to the intelligent system, it can then use the information to make decisions. Doing so exhausts this information, and system entropy increases to its original value. This, in turn, will drive the system to acquire new information in the next process cycle.

In its final phase, the intelligent system must physically prepare itself for the next cycle by ensuring it can record new actionable information. As consistent per Landauer's Principle [17] and the observations made by Bennett [18], this requires energy. More specifically, energy is required by the system to reset its memory, thereby allowing it to record new information. Landauer and Bennett state that this energy is used for memory erasure; however, this is not the perspective taken here. Instead, this energy is required for the system to transition back to its initial higher temperature state where it can then record new information. Information erasure is consequential.

3. Thermodynamic Computing

Given that intelligent processes are physical, one should be able to provide a more detailed account of their underlying physics. What emerges is an engineering formalism that is a hybrid combination of thermodynamics and information theory. Moreover, this formalism guides the way for the development of optimal and more complex intelligent processes. As will be shown, the required optimization can be solved either as a min-max Helmholtz free energy problem or as a min-max information optimization problem, and that these are equivalent problems. This has the engineering consequence that optimized computation and adaptation equate to optimized energy efficiency.

In either case of energy optimization or information optimization, the objective functions are convex in one of the optimization spaces and concave in the other, thus guaranteeing unique solutions at saddle points. For this reason, this section digresses in the area of convex optimization. This is believed critical to understanding how intelligent processes optimize their adaptability and how they operate.

Geometric Programming (GP) provides an efficient means of nonlinear optimization. It is thought to be critical in the practical development of practical operational algorithms and computer architectures for implementing optimal intelligent processes and communication systems. This is done through the context of double matching [19], as originally developed for optimizing communication systems. The relevance of GP to information theory is presented by Chiang [20], who describes its relevance to solving the basic source and channel coding problems of information theory. The duality and convexity properties of GP are well known. In essence, GP is a generalization of Linear Programming (LP) as extended to almost arbitrarily complex convex nonlinear optimization problems such as optimal source and channel coding and the joint optimization achieved through double matching. There is wide availability of efficient GP problem solvers including the interior-point algorithm [21–28]. Much of the following is a re-casting of Reference [20] in the context of optimizing intelligent processes.

Minimizing a convex function subject to inequality constraints and affine equality constraints means solving a problem of the form:

$$\begin{aligned} & \min f_0(\mathbf{x}) \\ & \text{subject to } f_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, m \\ & \quad A\mathbf{x} = \mathbf{c} \\ & \quad \mathbf{x} \in \mathbf{R}^n \end{aligned} \quad (1)$$

where $A \in \mathbf{R}^{n \times l}$ and $\mathbf{c} \in \mathbf{R}^l$ are constant parameters of the problem. The objective function to be minimized is f_0 , and the m constraint functions f_i are assumed convex. The objective function f_0 need not be convex.

From convexity analysis [24], it is well known that within convex optimization problems there exists a global minimum. Lagrange's dual theory means that optimization problems can be solved using either the primal problem or the Lagrange dual problem. The solution to the dual problem provides a lower bound to the solution of the primal (minimization) problem [24]. The optimal values of the primal and dual problems are not necessarily the same. The difference is called the duality gap. Under Slater's Condition [20,24], as is the case for entropic measures like mutual information, the duality gap is zero. Slater's Condition essentially states that a feasible solution exists given the posed inequality constraints.

GP is formulated as follows. First, there are two forms of GP optimization: *standard* and *convex*. In standard form, a constrained optimization is performed on a function called a posynomial. The convex form is obtained from the first through a logarithmic change of variable.

Posynomials are defined in terms of a sum of monomials. A monomial is a function $f(\mathbf{x})$ such that $f: \mathbf{R}_{++}^n \rightarrow \mathbf{R}^1$, where \mathbf{R}_{++}^n denotes the strictly positive quadrant of n -dimensional Euclidean space. A monomial is defined by:

$$f(\mathbf{x}) = d x_1^{a^{(1)}} x_2^{a^{(2)}} \dots x_n^{a^{(n)}} \quad (2)$$

where $d \geq 0$ is a multiplicative constant, as are the exponential constants $a^{(j)} \in \mathbf{R}, j = 1, 2, \dots, n$. A posynomial is then defined by a sum of monomials:

$$f(\mathbf{x}) = \sum_{k=1}^K d_k x_1^{a_k^{(1)}} x_2^{a_k^{(2)}} \dots x_n^{a_k^{(n)}} \quad (3)$$

where $d_k \geq 0, k = 1, 2, \dots, K$ and $a_k^{(j)} \in \mathbf{R}, j = 1, 2, \dots, n$. The important aspects of posynomials are their positivity and their convexity in convex form.

The GP problem in standard form corresponds to minimizing a posynomial subject to posynomial upper-bound inequality constraints and monomial equality constraints. Formally:

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{x}) \\ & \text{subject to} && f_i(\mathbf{x}) \leq 1, \quad i = 1, 2, \dots, m \\ & && h_l(\mathbf{x}) = 1, \quad l = 1, 2, \dots, M \end{aligned} \quad (4)$$

$$f(\mathbf{x}) = \sum_{k=1}^K d_k x_1^{a_k^{(1)}} x_2^{a_k^{(2)}} \dots x_n^{a_k^{(n)}}$$

where f_i are posynomials given by:

$$f_i(\mathbf{x}) = \sum_{k=1}^{K_i} d_{ik} x_1^{a_{ik}^{(1)}} x_2^{a_{ik}^{(2)}} \dots x_n^{a_{ik}^{(n)}} \quad (5)$$

and $h_l(\mathbf{x})$, $l = 1, 2, \dots, M$ are monomials $h_l(\mathbf{x}) = d_l x_1^{a_l^{(1)}} x_2^{a_l^{(2)}} \dots x_n^{a_l^{(n)}}$, and the minimization is over \mathbf{x} . Every GP problem in standard form can be represented by two data structures [20] denoted by the matrix A and the vector \mathbf{d} , and the mapping from these constants to their respective constraints and objective functions. A is a matrix of exponential constants, and \mathbf{d} is a vector of multiplicative constants.

Optimization of GP in standard form is not a convex optimization problem. However, it can be made so through a logarithmic change of variables and its multiplicative constants. In particular, using the transformations $y_i = \log x_i$, $b_{ik} = \log d_{ik}$, and $b_l = \log d_l$, a GP problem in standard form is transformed into the following GP convex optimization problem:

$$\begin{aligned} & \text{minimize} && \sum_{k=1}^K \exp[\mathbf{a}_{0k}^T \mathbf{y} + b_{0k}] \\ & \text{subject to} && \sum_{k=1}^{K_i} \exp[\mathbf{a}_{ik}^T \mathbf{y} + b_{ik}] \leq 1, \quad i = 1, 2, \dots, m \\ & && \mathbf{a}_l^T \mathbf{y} + b_l = 0, \quad l = 1, 2, \dots, M \end{aligned} \quad (6)$$

where $\mathbf{a}_{ik} = [a_{ik}^{(1)}, a_{ik}^{(2)}, \dots, a_{ik}^{(n)}]^T$ and the minimization is over \mathbf{y} .

Before proceeding, an important convexity relationship needs to be summarized. It is critical to understanding the fundamental tie between information theory and statistical mechanics [20]. An important lemma in GP is that the log-sum-exp function as defined by Equation (7) is convex in \mathbf{x} :

$$f(\mathbf{x}) = \log \sum_{i=1}^n e^{x_i} \quad (7)$$

This function has the functional form of the logarithm of the partition function Z of a Boltzmann distribution, which, as will be shown, is of importance. Proof of the convexity of Equation (7) can be seen through a simple inequality.

The log-sum inequality [29] is given by:

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^n a_i \right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \quad (8)$$

where each of the elements in \mathbf{a} and \mathbf{b} are greater than or equal to zero. For a function mapping \mathbf{R}^n to \mathbf{R} , there exists [30] a conjugate function f^* that is also a mapping from \mathbf{R}^n to \mathbf{R} and defined by:

$$f^*(\mathbf{y}) = \sup_{\mathbf{x}} (\mathbf{y}^T \mathbf{x} - f(\mathbf{x})) \quad (9)$$

The function f^* is guaranteed to be a convex function because it is the point-wise supremum of a family of affine functions of \mathbf{y} [31].

If $\hat{b}_i = \log b_i$ and $\sum_{i=1}^n a_i = 1$ in the inequality in Equation (8), then it follows that:

$$\log \left(\sum_{i=1}^n e^{\hat{b}_i} \right) \geq \mathbf{a}^T \hat{\mathbf{b}} + \sum_{i=1}^n a_i \log \frac{1}{a_i} \quad (10)$$

with equality only holding if and only if a_i has the form of the Boltzmann distribution, whereby:

$$a_i = \frac{e^{\hat{b}_i}}{\sum_j e^{\hat{b}_j}} \quad (11)$$

Because intelligent processes are characterized by Boltzmann distributions, as will be shown, this takes on a critical importance in understanding intelligent processes. To summarize, Equation (10) means that the log-sum-exp function is the conjugate function of entropy. The log-sum-exp function is convex because all conjugate functions are required to be so.

The practical importance of GP in convex form to optimizing intelligent processes arises from posing and solving the Lagrange dual problem, which is linearly constrained, and its objective function takes the form of a generalized entropy function. Using conventional Lagrange dual transformations as shown by Boyd [24], the GP convex optimization problem with m posynomial constraints becomes:

$$\begin{aligned} & \text{minimize } \mathbf{b}_0^T \mathbf{v}_0 - \sum_{j=1}^{K_0} v_{0j} \log v_{0j} + \sum_{i=1}^m \left[\mathbf{b}_i^T \mathbf{v}_i - \sum_{j=1}^{K_i} v_{ij} \log \frac{v_{ij}}{\mathbf{1}^T \mathbf{v}_i} \right] \\ & \text{subject to } v_i \geq 0, \quad i = 1, 2, \dots, m \\ & \quad \mathbf{1}^T \mathbf{v}_0 = 1 \\ & \quad \sum_{i=0}^m \mathbf{A}_i^T \mathbf{v}_i = 0 \end{aligned} \quad (12)$$

with the optimization performed over the vectors \mathbf{v}_i . Here, $\mathbf{1}$ is a column vector of all 1's. The length of \mathbf{v}_i is K_i , the number of monomial terms in the i th posynomial, $i = 1, 2, \dots, m$. The matrix \mathbf{A}_0 contains the exponential constants in the objective function with each row corresponding to the constituent monomials. \mathbf{A}_i , $i = 1, 2, \dots, m$ are the matrices of exponential constants in the m posynomial constraints. The multiplicative constants in the objective posynomial correspond to the vector \mathbf{b}_0 . Likewise, the \mathbf{b}_i matrices correspond to the multiplicative constants for the monomials defining the constraint posynomials.

There are several items of note in this formulation of the optimization problem. The dual variable \mathbf{v}_0 is a probability distribution. The other dual vectors \mathbf{v}_i are not necessarily so. The objective function consists of the sum of linear terms and entropy-like functions. The exponential constants of the monomials define a single linear equality constraint. Solutions for the GP problem in convex form are globally optimal, as are their solutions found through its Lagrange dual.

As previously discussed, information theory deals with two basic problems—those of source and channel coding. These, respectively, define the information-theoretic limits of data compression and its transmission rate through a channel. In the first problem, an average distortion or loss criteria is specified, and the communication system cannot exceed this level nor can the transmission be lowered further. In the latter, an average transmission cost is typically imposed below which the system must operate in the maximization of its transmission rate. Collectively, these two problems are known as Shannon's dual problems [13].

Within intelligent processes, the rate distortion problem becomes that of minimizing the rate at which that information is acquired and distilled (compressed) into actionable information. The second problem becomes that of maximizing the rate at which actionable information can be used to make

decisions and perform useful control. The duality of the problems of communication processes and intelligent processes can be simply understood from Bayes' Theorem. For a communication process with source x and receiver y , Bayes' Theorem for the joint density is factored as $p(x)p(y|x)$. This form is needed to address maximizing the source entropy $H(X)$ and minimizing $I(X; Y)$. Conversely, an intelligent process with acquired information X and control decision Y emphasizes the factorization $p(x, y) = p(x|y)p(y)$. This form addresses maximizing the decision rate $H(Y)$ while minimizing required information acquisition rate $I(X; Y)$ to achieve the maximized output. All of these optimization problems have a GP formulation. The practical implication is that communication and intelligent processes can be efficiently optimized.

Consider the problem of maximizing the *control authority* $H(Y)$ of an intelligent system subject to control costs for a discrete memoryless intelligent system like a cortical neuron. It is memoryless in the sense that the system response is only a function of the current input and that past inputs do not impact this decision except as through much longer time-scale self-adaptation. The term *control authority* describes the rate at which an intelligent process can make decisions and effect control. In the case of an animal, for example, it would represent the creature's physical agility or the accessible phase states of a physical system [32].

An intelligent process acquires information by asking $X = \{x_1, x_2, \dots, x_m\}$. The system then uses some portion of this information, the actionable information, to make decisions by answering $Y = \{y_1, y_2, \dots, y_q\}$. Maximizing control authority formally corresponds to finding:

$$C(S) = \max_{\{p(y_i)\}} I(X; Y) \quad (13)$$

where the optimization is performed over the output (decision) distribution $\{p(y_i)\}$.

Now define $r \in R^{q \times 1}$ with:

$$r_j = \sum_{i=1}^m p(x_i|y_j) \log \frac{1}{p(x_i|y_j)} = H(X|y_j) \quad (14)$$

then:

$$I(X; Y) = -\sum_{i=1}^m p(x_i) \log p(x_i) - p_y^T r \quad (15)$$

where p_y is a column vector of the probabilities $p(y_1), p(y_2), \dots, p(y_q)$. Likewise, define the column vector p_x of probabilities $p(x_i)$. When expressed in this form, it is clear that intelligent process capacity optimization is the Lagrange dual of a convex GP problem. Including an average decision cost constraint of the form $\sum_{j=1}^q s_j p(y_j) \leq S$, the formal Lagrange dual formulation of capacity maximization problem is given by:

$$\begin{aligned} & \text{maximize} \quad -p_y^T r - \sum_{i=1}^m p(x_i) \log p(x_i) \\ & \text{subject to} \quad P^T p_y = p_x \\ & \quad \quad p_y^T s \leq S \\ & \quad \quad \mathbf{1}^T p_y = 1 \\ & \quad \quad p(y_i) \geq 0 \end{aligned} \quad (16)$$

where the optimization is carried out over p_y and p_x for fixed P . The matrix P is the m by q of conditional probabilities $p(x_i|y_j)$. This problem can equivalently be expressed in convex form as:

$$\begin{aligned} & \text{minimize} \quad \log \sum_{j=1}^q \exp(\alpha_j + \gamma S) \\ & \text{subject to} \quad P\alpha + \gamma s \geq -r \\ & \quad \quad \gamma \geq 0 \end{aligned} \quad (17)$$

where the optimization is carried out over α and γ , and where P , s , and S are constants.

The dual problem to optimizing the control capacity of an intelligent system is that of minimizing its control rate subject to a maximum control error. This problem is the information-theoretic analog of minimizing the transmission rate through a communications channel subject to a reconstruction distortion constraint D . Formally, to minimize the control rate, the objective is to:

$$R(D) = \min_{\{p(x_i|y_j)\}} I(X;Y) \quad (18)$$

subject to the constraint:

$$\sum_{j=1}^q \sum_{i=1}^m p(y_j) p(x_i|y_j) d_{ij} \leq D \quad (19)$$

where d_{ij} is the cost of making decision y_j given it observes x_i . Regarding the cortical neural model (described further later), its input X consists of 2^n possible codes, where $n \approx 10,000$. During learning, the neuron partitions this space of codes into two sets, X_0 and X_1 . As the temperature T goes to zero, the neuron deterministically decides to not fire if it observes a code in X_0 and, conversely, to fire if an observed code belongs to X_1 . Note that $X_0 \cap X_1 = \phi$, the empty set, and $X_0 \cup X_1 = X$. However, with the neural computational temperature taking the form of an equivalent noise [33], the neuron will make errors and thus wrongly decide to fire when it should not and to not fire when it should. Define (1) the control distortion measure $d_{ij} = 1$ where $d_{ij} = 1$ if $i = j$ for observances from X_i , $i = \{0, 1\}$; and (2) the binary firing decisions y_j , where $j = \{0, 1\}$. The D in Equation (19) is the control error rate. For the cortical neuron, $0 \leq D \leq 1$ bits per decision are its possible transmission rates. Moreover, it is the intelligence rate of the process and is regulated by the temperature of the system during the time it makes a decision. Lastly, D corresponds to the Carnot or operational energy efficiency of the device.

Returning to the control rate minimization of Equations (18) and (19), this problem [20] is stated as:

$$\begin{aligned} & \text{minimize } \sum_{i=1}^m \sum_{j=1}^q p(y_j) p(x_i|y_j) \log \frac{p(x_i|y_j)}{\sum_{k=1}^q p(y_k) p(x_i|y_k)} \\ & \text{subject to } \sum_{i=1}^m \sum_j p(y_j) p(x_i|y_j) d_{ij} \leq D \\ & \sum_{i=1}^m p(x_i|y_j) = 1 \quad \text{for } j = 1, 2, \dots, q \\ & p(x_i|y_j) \geq 0 \end{aligned} \quad (20)$$

with the optimization carried out with respect to the conditional probabilities $p(x_i|y_j)$ that define the matrix P . The problem in Equation (20) can be expressed as:

$$\begin{aligned} & \text{minimize } -\mathbf{p}_y^T \mathbf{r} - \sum_{i=1}^m p(x_i) \log p(x_i) \\ & \text{subject to } \mathbf{P}^T \mathbf{p}_y = \mathbf{p}_x \\ & \sum_{i=1}^m \sum_{j=1}^q p(y_j) p(x_i|y_j) d_{ij} \leq D \\ & \sum_i p(x_i|y_j) = 1 \quad \text{for } j = 1, 2, \dots, q \\ & p(x_i|y_j) \geq 0 \end{aligned} \quad (21)$$

with the optimization performed with respect to \mathbf{P} with fixed \mathbf{p}_y . Keeping in mind the definition for \mathbf{r} , this can be expressed as:

$$\begin{aligned} & \text{minimize } -\mathbf{p}_y^T \mathbf{H}(X|y) - \sum_{i=1}^m p(x_i) \log p(x_i) \\ & \text{subject to } \mathbf{P}^T \mathbf{p}_y = \mathbf{p}_x \\ & \quad \sum_{i=1}^m \sum_{j=1}^q p(y_j) p(x_i|y_j) d_{ij} \leq D \\ & \quad \sum_i p(x_i|y_j) = 1 \text{ for } j = 1, 2, \dots, q \\ & \quad p(x_i|y_j) \geq 0 \end{aligned} \quad (22)$$

where $\mathbf{H}(X|y)$ is a column vector of length q of the conditional entropies $H(X|y_i)$, $i = 1, 2, \dots, q$. The convex GP formulation [20] for minimizing $R(D)$ is given by:

$$\begin{aligned} & \text{maximize } \mathbf{p}_y^T \boldsymbol{\alpha} - \gamma D \\ & \text{subject to } \log \sum_{i=1}^q \exp(\log p_i + \alpha_i - \gamma d_{ij}) \\ & \quad \gamma \geq 0 \end{aligned} \quad (23)$$

with the optimization carried out over $\boldsymbol{\alpha}$ and γ . The fact that this problem is an energy minimization problem was first recognized by Berger [34].

Control authority maximization and control rate minimization both contain the log-sum-exp function either as the function to be optimized (Equation (17)) or as a constraint (Equation (23)) for GP in convex form. The log-sum-exp function is the log of the partition functions for Boltzmann distributions in statistical physics. The implication is that the rate and channel coding in information theory and the control authority and control rate problems associated with intelligent processes can both be interpreted from this perspective.

Consider a physical system containing n states. Each state i has a corresponding energy e_i and probability p_i . Define the n -vectors for energy \mathbf{e} defined by e_i and \mathbf{p} defined by p_i . The average energy of the system is given by $U(\mathbf{p}, \mathbf{e}) = \mathbf{p}^T \mathbf{e}$. Its entropy is determined by $H(\mathbf{p}) = -\sum p_i \log p_i$. The Helmholtz free energy [20] of the system is given by:

$$F(\mathbf{p}, \mathbf{e}) = U(\mathbf{p}, \mathbf{e}) - TH(\mathbf{p}) \quad (24)$$

Solving for the minimum Helmholtz free energy of the system over \mathbf{p} requires solving:

$$\begin{aligned} & \text{minimize } \mathbf{p}^T \mathbf{e} + T \sum_{i=1}^n p_i \log \frac{1}{p_i} \\ & \text{subject to } \mathbf{1}^T \mathbf{p} = 1, \mathbf{p} \geq 0 \end{aligned} \quad (25)$$

whereas its corresponding GP problem in convex form requires maximizing the partition function of the system. Note that if one identifies energy e_i with $-r_i$ in Equation (21), one has exact correspondence of maximizing Helmholtz free energy and the control authority C of an intelligent process. Similarly, Helmholtz free energies constrain the control rate minimization problem of Equation (23).

Because $F(\mathbf{p}, \mathbf{e})$ is convex in \mathbf{p} and concave in \mathbf{e} , it is true that mutual information is the same for $\{p(y_j)\}$ and $\{p(x_i|y_j)\}$, respectively. This means that:

$$\max_{\mathbf{p}} \min_{\mathbf{e}} F(\mathbf{p}, \mathbf{e}) = \min_{\mathbf{e}} \max_{\mathbf{p}} F(\mathbf{p}, \mathbf{e}) \quad (26)$$

This joint optimization problem thermodynamically corresponds to maximizing Helmholtz free energy for the worst energy assignment. From an information-theoretic perspective, it is minimizing channel throughput and matching it to the source entropy rate. This corresponds

to double-matching [19]. For an intelligent process, Equation (26) also indicates this minimizing acquisition rate, thus maximizing control authority, and the dynamic matching of the two objectives is dynamically achieved through learning and adaption. Thus, learning serves to make the intelligent process simultaneously pursue optimal information processing and optimal energy efficiency. This is most greatly demonstrated in the operation of ostensibly all biological systems and especially so in brains.

As will be discussed, the proposed optimal adaptation process does not pursue the optimization in Equation (26) simultaneously. Rather, one can alternate repeatedly between minimizing F over e and maximizing it over p . Convergence to global optima are guaranteed by the convexity and concavity of F with respect to p and e , respectively.

4. Reverse-Engineering the Cortical Neuron

This section presents eight topics:

- Overview
- Logical structure of a cortical neuron
- Boltzmann distribution and neural architecture
- Double-matching and spatiotemporal adaptation
- Neural operation
- A training example
- Thermodynamic computing perspective

4.1. Overview

Cortical neurons as found within the human brain have on the order of 10,000 dendritic inputs and a single output. Inputs and outputs of each neuron consist of either no signal or an action potential generated at some time. Action potentials are the embodiment of the notion of an elementary assertion. They either exist or do not at any point in time.

Each input dendrite to a neuron allows it to ask a question $X_i \equiv$ "Is there an observed pre-synaptic potential or not, and therefore $X_i \equiv \{x_i, \neg x_i\}$?" Like the photon detector, the dendrite need only physically ask either of the elementary questions, $X_i = \overline{x_i}$ or $\neg X_i = \overline{\neg x_i}$ and like the photon detector, choose the first detection strategy in asserting the positive when a photon arrives.

The solitary output of each cortical neuron corresponds to an answer to a question $Y \equiv$ "Should I fire or not?" This question is formally defined by $Y \equiv \{y, \neg y\}$. Physically, it need only answer either assertion to convey control to its environment, and of course, biological neurons choose to answer the assertion by generating an action potential $Y = \overline{y_i}$ or not.

After establishing the basic relationship between biological signals and the logical formalism of questions and assertions, one can proceed to understanding the logical architecture of the cortical neuron and then how they process information and operate as highly efficient TCs.

4.2. Logical Structure of a Cortical Neuron

Logical questions provide a simple and elegant way of capturing the computational nature of cortical neuron from the "first person" perspective in terms of what it can "see" and what it can "do".

Consider first the information that the neuron can acquire from its neural environment, and in doing so, first look at a single dendrite. Each dendrite asks a binary question $X_i \equiv \{x_i, \neg x_i\}$, with x_i corresponding to the assertion of a presynaptic action potential or the lack of the detection of such. A typical human cortical neuron has on the order of $n = 10,000$ dendrites, a number likely limited by physiological constraints. Collectively, the dendritic field allows the neuron to ask $X = X_1 \wedge X_2 \wedge \dots \wedge X_n$; a question defined by $N = 2^n$ assertions corresponding to all of the codes the neuron can possibly observe. N is a huge number corresponding to all possible input neural microstates; thus, the thermodynamic nature of a single cortical neuron.

The neuron is capable of answering a single binary question, $Y \equiv \{y, \neg y\}$. This decision is made at the axon hillock [35] of the cortical neuron. Two possible output decision states, combined with 2^n informational states, means that there are a total of 2^{n+1} assertions that define possible microstates of the neuron. These assertions define the joint question $X \wedge Y$. It is important to realize that these states are only defined internally and subjectively by the neuron.

The neuron uses the actionable information defined by $X \vee Y$ to make its decisions Y . It follows that:

$$\begin{aligned} X \vee Y &= (X_1 \wedge X_2 \wedge \dots \wedge X_n) \vee Y \\ &= (X_1 \vee Y) \wedge (X_2 \vee Y) \wedge \dots \wedge (X_n \vee Y) \end{aligned} \quad (27)$$

Upon finding the reflections of the expressions in Equation (27), one obtains:

$$\overline{X \vee Y} = (x_1 \wedge y) \vee (x_2 \wedge y) \vee \dots \vee (x_n \wedge y) \quad (28)$$

The significance of this expression is as follows: Conjunction “ \wedge ” has the physical significance of “coincidence detector” as the meaning of $X \vee Y$. Conversely, disjunction has the physical significance of superposition. Thus, the expression in Equation (28) means that the neuron is observing the coincidences between each dendritic assertion and its output assertion y . Furthermore, it finds the superposition of all observed coincidence events.

Lastly, suppose that the neuron can observe its own output decision y by asking Y . The composite question posed by the cortical neuron then becomes $(X \vee Y) \wedge Y$. One can note that:

$$\overline{(X \vee Y) \wedge Y} = [(x_1 \wedge y) \vee (x_2 \wedge y) \vee \dots \vee (x_n \wedge y)] \vee y \quad (29)$$

which captures the overall logical architecture of the cortical neuron.

4.3. Boltzmann Distribution and Neural Architecture

The logical expression in Equation (29) suggests that the cortical neuron is able to observe the n coincidences $x_i y$ and its output y . Moreover, suppose that it can observe the moments $E\{x_i y\} = \langle x_i y \rangle$, $i = 1, 2, \dots, n$, and $E\{y\} = \langle y \rangle$. Then, as in thermodynamics, one can apply the maximum entropy principle [36] to obtain the single cortical neuron Boltzmann distribution given by:

$$\begin{aligned} p(x, y) &= \frac{\exp\left[-\left(\sum_{i=1}^n \lambda_i x_i y - \mu y\right)/T\right]}{Z} \\ &= \frac{\exp[-e_i/T]}{Z} \end{aligned} \quad (30)$$

where Z is the partition function given by:

$$\sum_{y' \in Y} \sum_{x' \in X} \exp\left[-\left(\sum_{i=1}^n \lambda_i x'_i y' - \mu y'\right)/T\right] \quad (31)$$

As will be shown, the n -vector Lagrange λ corresponds to the dendritic efficacies or gains, whereas the Lagrange factor μ corresponds to the neural decision threshold.

Almost every marginal and conditional probability distribution of Equation (30) conveys insight in the probabilistic nature of the cortical neuron. Two distributions of particular interest are its *a posteriori* log-odds given by $\log p(y = 1 | x) / p(y = 0 | x)$ and the conditional probability $p(y = 1 | x)$.

The log odds [36] is given by:

$$\begin{aligned} \log \frac{p(y=1|x)}{p(y=0|x)} &= \lambda^T x - \mu \\ &= \log \frac{p(x|y=1)}{p(x|y=0)} + \log \frac{p(y=1)}{p(y=0)} \\ &= v \end{aligned} \quad (32)$$

The conditional probability $p(y = 1 | x)$ is given by:

$$p(y = 1 | x) = \frac{1}{1 + \exp[-v/T]} \quad (33)$$

The implications of Equations (32) and (33) are significant. The soma is ostensibly a spatiotemporal integrator with a time constant τ that is on the order of milliseconds of precision. Equation (32) means that the soma can integrate weighted inputs and then compare this integral with a decision threshold μ and implement Bayes' Theorem and guide its decisions in this way. Equation (33) implies that it can use this numeric quantity directly to make the probabilistic decision to fire $y = 1$ or not $y = 0$ as performed at the axon hillock.

The conditional probabilities $p(x_i = 0 | y = 1)$ and $p(x_i = 1 | y = 1)$ also are interesting. From Equation (30) it can be seen that:

$$p(x_i | y = 0) = \frac{1}{2} \quad (34)$$

and furthermore that $p(x | y = 0) = 2^{-n}$. This means that $H(X | y = 0) = n$ bits, which means it has maximal uncertainty of what its x input is when it does not fire. The probability $p(x_i = 1 | y = 1)$ is given by:

$$p(x_i = 1 | y = 1) = \frac{1}{1 + \exp(-\lambda_i)} \quad (35)$$

In Section 4.4, the requirement that the L^2 -norm $|\lambda|^2 \leq 1$ is enforced. Because $n \approx 10,000$, all typical connection strengths $|\lambda_i|$ are very small and close to zero with regard to Equation (35). Therefore, all $p(x_i | y = 1)$ can be expected to deviate only modestly from $1/2$. Consequentially, $H(X_i | y = 1)$ will be slightly less than 1 with the difference $1 - H(X_i | y = 1)$ being the small quantity of information contained in any single dendritic input regarding the output decision. The sum of these n differences across the dendritic field X is the total actionable information $I(X; Y)$ in making its decisions Y .

As will be described, the requirement of the simultaneity of the presentation of stimuli to the soma as required by Equation (28) is not arbitrary, but rather is an achieved condition. The optimal Hebbian [37] adaptation rules for the n -vector λ and decision threshold μ will be seen to extend to the requirement for the adaptation of a dendritic delay equalization n -vector τ defined by learned delay times τ_i between the input of an assertion into the i th dendrite and its integrated effect $\lambda_i x_i$ within the soma. Thus, given dendritic gains λ , decision threshold μ , and delay equalization vector τ , there are a total of 2^{n+1} adaptable parameters in this cortical model.

4.4. Double-Matching and Spatiotemporal Adaptation

The proposed neural optimization strategy of the neuron is to maximize its output decision rate while minimizing and equaling it to the rate at which it acquires information from its neural environment. This is the double-matching problem of information theory [19], which for the cortical neuron becomes:

$$\begin{aligned} & \min_{\lambda} I(X; Y) \\ & \text{subject to } |\lambda|^2 = \alpha \\ & \quad E\{d(x, y)\} \leq D \end{aligned} \quad (36)$$

and:

$$\begin{aligned} & \max_{\mu} I(X; Y) \\ & \quad E\{c(y)\} \leq C \end{aligned} \quad (37)$$

Equation (36) includes a constraint on the size of the L^2 norm of the dendritic n -vector. In general, $\alpha \leq 1$; therefore, in general $|\lambda_i| \ll 1$ given $n \approx 10,000$. This minimization problem can be solved whereupon the distortion constraint can be introduced and achieved by regulating the temperature

range $[T_l, T_u]$ over which the neuron operates in its Carnot cycle. This in effect regulates the Carnot efficiency of the neuron to D . Temperature regulation will be seen to correspond to noise regulation within the soma as controlled by the probability of error within each synaptic channel. This is accomplished via each dendritic channel corresponding to a binary erasure channel [38], whereby presynaptic potentials are not detected post-synaptically. This phenomena is known as Quantum Synaptic Failure or QSF [38].

Regarding the output optimization in Equation (37), no explicit decision cost is imposed. The presence of this constraint would serve to limit the maximum decision rate $H(Y)$ in bits per second to which the cortical neuron can operate. Within a biological neuron, this constraint must be present. One can imagine that if constructed, artificial cortical neurons could operate optically, electrically, or via other means that would allow them to operate at higher speeds than biological neurons.

Because the optimization problems in Equations (36) and (37) are convex, one merely needs to find the partial derivatives in each case, with respect to λ_i in the first case and to μ in the second, and then set them to zero in each case. The math is simplified through the use of what Fry refers to as the Gibbs Mutual Information Theorem [36] and more properly referred to in this paper as the Boltzmann Mutual Information Theorem. This theorem was independently derived by Hinton and Sejnowski [39].

As shown by Fry [36], if:

$$p(\mathbf{x}, \mathbf{y}) = \frac{\exp[-\boldsymbol{\theta}^T \mathbf{f}(\mathbf{x}, \mathbf{y})]}{Z} \quad (38)$$

then:

$$\frac{\partial I}{\partial \theta_i} = E \left\{ [f_i(\mathbf{x}, \mathbf{y}) - \langle f_i(\mathbf{x}, \mathbf{y}) \rangle] \log \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} \right\} \quad (39)$$

Regarding the dendritic strengths, the optimal λ in [36] is given by the largest eigenvector of the covariance matrix:

$$R = \langle \mathbf{x} \mathbf{x}^T \mathbf{y} \rangle - \frac{\langle \mathbf{x} \mathbf{y} \rangle \langle \mathbf{x} \mathbf{y} \rangle^T}{\langle \mathbf{y} \rangle} \quad (40)$$

where:

$$R\lambda = -\alpha\lambda \quad (41)$$

Solving for the optimal decision threshold μ gives:

$$\mu = \lambda^T E\{\mathbf{x}|\mathbf{y} = 1\} \quad (42)$$

The optimal λ can be found using a modified Hebbian form of E. Oja's equation [40]; thus, conditioned on $\mathbf{y} = 1$, the following equation is used to determine λ :

$$\lambda(t + \Delta t) = \lambda(t) + \pi v(t) [x(t) - \alpha^{-2} v(t) \lambda(t)] \quad (43)$$

where, per Equation (32), $v(t) = v[\mathbf{x}(t)]$ and π is a time constant of adaptation. Thus, a biologically plausible mechanism exists for performing the optimization in Equation (36). Optimizing Equation (37) is simple in that all that is needed is a time-averaging of the induced somatic potential in Equation (42) conditioned on $\mathbf{y} = 1$.

Presynaptic action potentials impinging on a dendrite can be expected to have variability inter-arrival times relative to the time that the neuron makes a decision. These potentials, although perhaps conveying actionable information to decisions being made, may arrive too early or too late to impact decisions. One can postulate that there exists a third Hebbian adaptation mechanism whereby the time delay τ_i comes from the detection of an action potential, its propagation along a dendrite, and the superposition of its effect on the somatic potential. Indeed, this is the case.

As shown by Fry [41], the following intuitive adaptation rule serves to optimally equalize the dendritic delays contained in the n -vector τ :

$$\frac{d\tau_i}{dt} = -\beta\lambda_i y(t) dx_i(t - \tau_i) / dt \quad (44)$$

The first point to note in this rule is that it is proportional to the connection strength λ_i , meaning that stronger connections expedite delay equalization. Secondly, this rule is like the other two rules, also Hebbian with its execution dependent on $y(t) = 1$ at reference decision time t . Lastly, depending on the sign of λ_i , the derivative can either be negative or positive, and the signs of the derivative and the synaptic efficacy cancel each other out, which guarantees the convergence of the delay τ_i to yield maximal synchrony of the arrival of dendritic effects at the soma for decisioning when $y = 1$.

Figure 5 graphically depicts the delay equalization process. The implication is that neural pathways communicate spatial-temporal population codes.

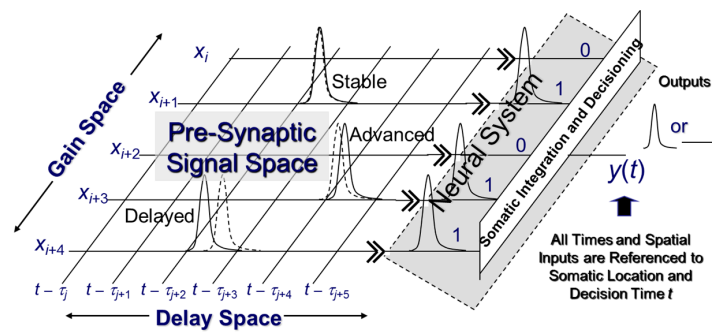


Figure 5. Depiction of delay equalization process relative to neural decisions $y(t)$ made at reference time t .

4.5. Neural Operation

Neural decisions are probabilistic and made according to Equation (33), which is a function of temperature T . This function is the logistic function of statistical regression. Let $\beta = 1/T$ be inverse temperature. Then, as shown in [42], $p(y = 1 | v + \eta) = \text{erf}[v / (2^{1/2} \sigma_n)]$ closely approximates $p(y = 1 | v)$ if η is zero-mean Gaussian noise with variance $\sigma_n^2 = 4e^{-2\beta/\sqrt{2\pi}}$. An overlay of these two functions as a function of v is provided in Figure 6.

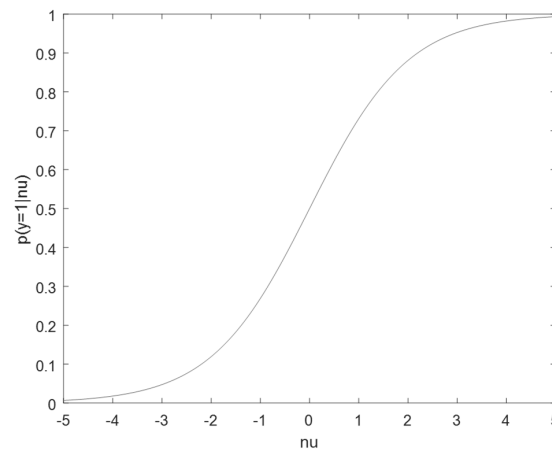


Figure 6. Overlay of the logistic function and error functions for $T = \beta = 1$ showing they are indistinguishable.

The implication is that by combining additive white noise to the induced somatic potential, one can implement a probabilistic decision rule consistent with Equation (33). Probabilistic decisioning with errors is a requirement for adaptation to occur. As described in Section 4.4, QSF provides a simple means of generating Gaussian noise. The ideal potential induced in the soma is given by:

$$\nu = \sum_{i=1}^n \lambda_i x_i \quad (45)$$

Through QSF, some fraction of the input potentials will randomly be eliminated such that the noisy somatic potential becomes:

$$\begin{aligned} \zeta &= \sum_{i=1}^n \lambda_i x_i - \sum_{j \in Q} \lambda_j x_j \\ &= \nu + \eta \end{aligned} \quad (46)$$

where Q is the random set of failed synaptic junction events. Because connection strengths are randomly positive and negative and the dendrites are assumed mutually independent of one another, the sum of inputs of failed junctions has a Gaussian distribution. The central limit theorem holds because QSF occurs 50% to 75% of the time [38]. The QSF rate essentially determines the operating temperature of the neuron as well as its achievable Carnot efficiency, which seems to range between 25% and 50%, which is significant if such efficiencies are possible. Such efficiencies offered by cortical neurons may contribute significantly to the overall efficiencies of brains.

4.6. A Training Example

Suppose one has 20 codes each containing a 20-bit code. These codes are delineated in Figure 7. Some codes are replicated, whereas others not.

		Bit																			
Code Index		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	1	1	1	1	0	1	1	0	1	0	0	0	1	1	0	1	1	0	0	1
	2	0	1	0	0	0	1	0	0	0	0	0	1	1	1	0	1	1	1	0	1
	3	1	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1
	4	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	0	1
	5	1	1	1	1	0	1	1	0	1	0	0	0	1	1	0	1	1	0	0	1
	6	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1
	7	0	1	0	1	0	1	1	0	1	0	0	0	0	1	0	1	1	0	0	1
	8	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	0	1
	9	1	1	1	1	0	1	1	0	1	0	0	0	1	1	0	1	1	0	0	1
	10	1	0	0	1	0	0	0	1	1	0	1	1	1	1	0	1	1	0	0	1
	11	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	1	0	1	1
	12	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	0	1
	13	1	1	1	1	0	1	1	0	1	0	0	0	1	1	0	1	1	0	0	1
	14	1	0	0	0	1	1	1	0	1	0	1	1	0	0	1	1	1	1	0	1
	15	0	0	0	0	1	1	1	1	0	0	1	0	1	0	1	0	0	1	1	1
	16	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	0	1
	17	0	1	1	0	1	1	1	0	0	0	1	1	1	1	0	1	0	0	1	1
	18	1	0	0	0	0	0	1	0	0	0	1	0	1	1	1	1	0	1	1	1
	19	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	1	1
	20	1	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1

Figure 7. Sample training codes used against the specified neural optimization objectives.

Figure 8 captures code similarities by tabulating the Hamming distances between the selected codes. For instance, one can see that the first code has four instantiations given it has a Hamming distance of zero from the codes indexed by itself, 1, and codes 5, 9, and 13.

Figure 9a illustrates which codes illicit action potentials $y = 1$ with probabilities greater than half. In particular, 9 of the 20 codes fall into this category. Ideally, should the training set allow, the set of input codes is split equally between those that induce an action potential and those that do not. This maximizes the output entropy $H(Y)$ of the cortical neuron by letting it approach 1 bit per decision.

Figure 9b plots the dendritic gains λ_i vs. index i . These values are equally likely to be positive or negative. Furthermore, repeated learning trials yield the same values; however, the sign of the vector λ varies in a random fashion between positive and negative. In opposing cases, the neuron fired on complementary code sets.

		Code Number																			
Code Number		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	0	7	10	1	0	7	3	1	0	7	6	1	0	10	20	1	8	10	12	7
	2	7	0	11	8	7	12	6	8	7	8	7	8	7	9	13	8	7	9	9	10
	3	10	11	0	9	10	7	9	9	10	9	8	9	10	12	10	9	14	10	6	5
	4	1	8	9	0	1	6	4	0	1	8	7	0	1	9	19	0	9	11	11	6
	5	0	7	10	1	0	7	3	1	0	7	6	1	0	10	20	1	8	10	12	7
	6	7	12	7	6	7	0	8	6	7	10	7	6	7	9	13	6	9	11	9	6
	7	3	6	9	4	3	8	0	4	3	8	5	4	3	9	17	4	9	11	11	6
	8	1	8	9	0	1	6	4	0	1	8	7	0	1	9	19	0	9	11	11	6
	9	0	7	10	1	0	7	3	1	0	7	6	1	0	10	20	1	8	10	12	7
	10	7	8	9	8	7	10	8	8	7	0	11	8	7	9	13	8	11	9	7	8
	11	6	7	8	7	6	7	5	7	6	11	0	7	6	10	14	7	8	8	8	7
	12	1	8	9	0	1	6	4	0	1	8	7	0	1	9	19	0	9	11	11	6
	13	0	7	10	1	0	7	3	1	0	7	6	1	0	10	20	1	8	10	12	7
	14	10	9	12	9	10	9	9	9	10	9	10	9	10	0	10	9	10	8	10	9
	15	20	13	10	19	20	13	17	19	20	13	14	19	20	10	0	19	12	10	8	13
	16	1	8	9	0	1	6	4	0	1	8	7	0	1	9	19	0	9	11	11	6
	17	8	7	14	9	8	9	9	9	8	11	8	9	8	10	12	9	0	8	14	11
	18	10	9	10	11	10	11	11	11	10	9	8	11	10	8	10	11	8	0	10	11
	19	12	9	6	11	12	9	11	11	12	7	8	11	12	10	8	11	14	10	0	9
	20	7	10	5	6	7	6	6	6	7	8	7	6	7	9	13	6	11	11	9	0

Figure 8. Hamming distances between selected training codes.

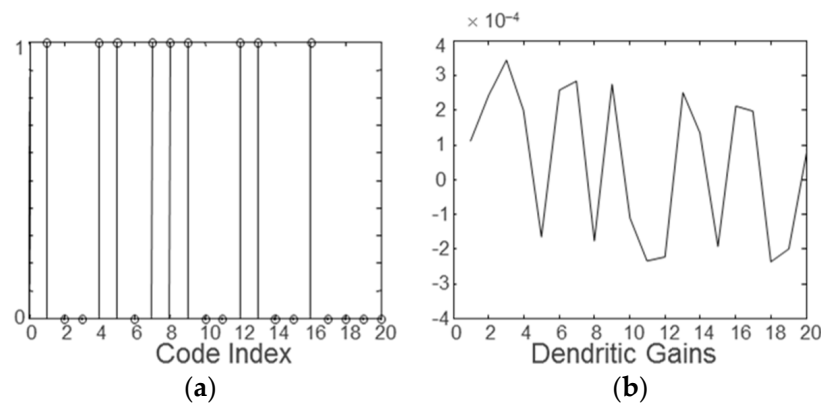


Figure 9. (a) Codes that illicit neural firing and (b) the depiction of dendritic field connection strengths.

Neural adaptation rules can be thought of as geometric. The quantity $\lambda^T x - \mu$ is the log *a posteriori* odds as per Equation (32) with $\mu = \log p(y=0)/p(y=1)$ and $\lambda^T x = \log p(x|y=1)/p(x|y=0)$. As depicted in Figure 10, the optimally adapted λ and μ define a hyper-plane. The vector λ defines the orientation of the hyper-plane, whereas μ defines its offset from the origin.

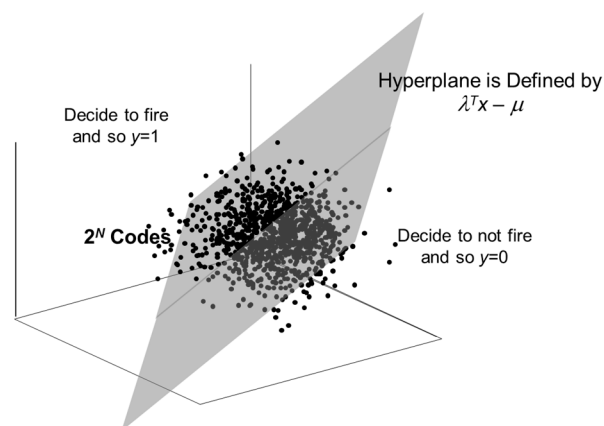


Figure 10. Hyper-plane representation for double-matching optimization.

Optimal hyper-planes slice through the n -dimensional space in such a way as to ensure the maximal dispersion of observed codes about each of its sides. The offset μ is such so as to ensure this dispersion has equal probabilities of inducing an action potential or not doing so.

4.7. Thermodynamic Computing Perspective

If one has a closed-form expression for the partition function of a system, then one knows basically everything thermodynamically about the system. The partition function for this simple cortical model is no different.

The partition function for the cortical neuron was given in Equation (31) and is repeated here using inverse-temperature β :

$$Z = \sum_{y \in Y} \sum_{x \in X} \exp \left[-\beta \left(\sum_{i=1}^n \lambda_i x_i y - \mu y \right) \right] \quad (47)$$

Summing over y gives:

$$Z = 2^n + 2^n e^{-\mu} \sum_{x \in X} \exp \left[-\lambda^T x \right] \quad (48)$$

Note that the optimal decision threshold is:

$$\mu = E \left\{ \lambda^T x \mid y = 1 \right\} \quad (49)$$

and that using the known form for the probabilities gives:

$$\begin{aligned} \mu &= \sum_{i=1}^n \lambda_i p(x_i = 1 \mid y = 1) \\ &= \sum_{i=1}^n \lambda_i \frac{1}{1 + e^{-\lambda_i}} \end{aligned} \quad (50)$$

After finding the Taylor expansion of each probability $p(x_i = 1 \mid y = 1) = \frac{1}{1 + e^{-\lambda_i}}$ about $\lambda_i = 0$, keeping the first two terms, and finally noting that:

$$\sum_{x \in X} \exp \left(-\lambda^T x \right) = \prod_{i=1}^n \left(1 + e^{-\lambda_i} \right) \quad (51)$$

yields the cortical partition function:

$$\begin{aligned} Z(n, \beta) &= 2^n + 2^2 e^{-\beta^2/4} \prod_{i=1}^n \cosh \left(\frac{\beta \lambda_i}{2} \right) \\ &= 2^n + Z_1 \end{aligned} \quad (52)$$

Z is plotted as a function of β and n in Figure 11. The function Z_1 reflects dependency on β , while both terms depend on n . As the number of cortical inputs n increases, one can see that the neuron can operate at successive lower temperatures, meaning that its energy efficiency increases with the number of input connections n . Thus, physiological constraints on this number is important and therefore it can be seen that optimal computational efficiency asymptotes at observed biological dendritic field sizes of $n \approx 10,000$. The deep red region shown in Figure 11 is where Z can take on the value 2^{n+1} . This is the maximum value it can have and is consistent with the fact that the system desires to maximize its Helmholtz free energy in the sense of Equations (23) and (26).

As suggested by Figure 11, the intelligent Carnot process executed by the cortical neuron cycles between two temperatures, $T = 1$ and $T = 0.1$, or as dictated by the lower temperature limit of the operation.

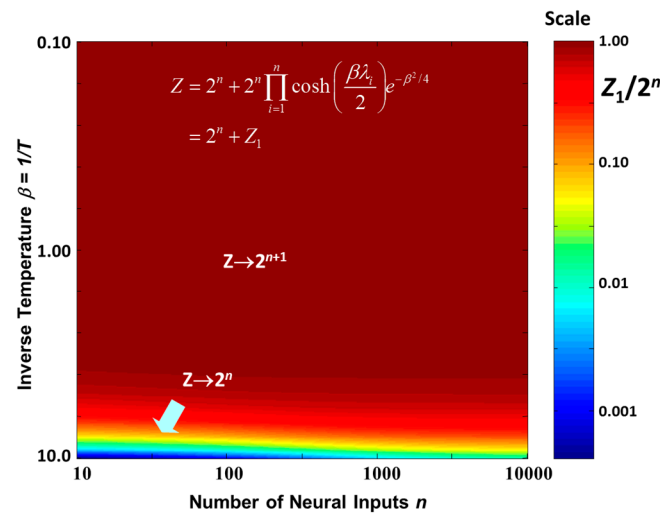


Figure 11. Colorized map of $Z(n, \beta)$ showing where phase transition occur. The dark red region is required for cortical operation.

All of this suggests that the cortical neuron operates according the Carnot process depicted in Figure 12. In Figure 12, the entropy of the neuron decreases, owing to the measurement of a new dendritic field potential x in phase 1. During phase 2, the actionable information or potential is deposited within the soma, less the effects of Gaussian noise introduced through QSF. Phase 2 is also associated with a decrease in neural temperature, meaning that the induced noisy potential has been recorded and stored within the soma where it is no longer subject to the effects of noise. During phase 3 of the intelligent Carnot process, the neuron decides whether to fire or not, and in either case, expends and dissipates collected information. As actionable information is expended in decisioning, system entropy increases. Lastly, during the fourth and final phase, the neuron reestablishes sodium and potassium ion concentration biases across the cell membrane. This allows the cortical neuron the renewed ability to acquire new information and make new decisions as to whether to fire or not.

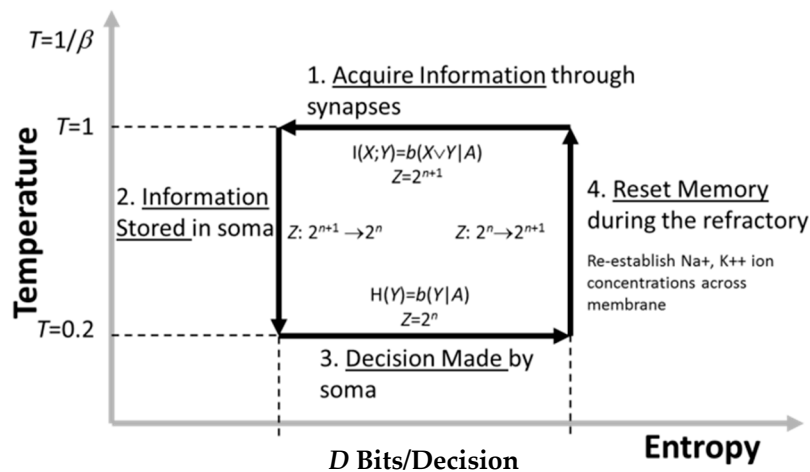


Figure 12. Depiction of Carnot process underlying the operation of the cortical neuron.

During temperature decreases and increases in the process cycle shown in Figure 12, the neural system undergoes phase state changes whereby Z is reduced from 2^{n+1} to 2^n and then restored to 2^{n+1} . During the first cycle, actionable information in the form of the sensed somatic potential is stored, thus allowing the neuron to make one decision or another. During the latter cycle, the neuron's ability to make this determination is restored. In this phase in the endothermic phase of cortical operation,

energy resources in the form of adenosine triphosphate energize sodium-potassium ion pumps to restore ion concentrations across the cortical cell membrane.

Z operationally transitions between 2^n and 2^{n+1} . Until a neural decision is made, the neuron has 2^{n+1} states, including the two possible output decision states. After firing, and when the neuron enters its refractory state, the output state is no longer accessible and Z transitions to 2^n . Physically, this corresponds to the neuron undergoing phase transitions, whereby the decision state y is accessible or not.

5. Discussion

This paper introduced a physics-based explanation of intelligent processes and the idea of thermodynamic computing. There are many additional considerations that go well beyond those considered. For example, because energy is an extensive property of physical systems (Extensive is meant to mean the narrow interpretation that the property scales with the size of the system and not for instance as applies to small-system thermodynamics as described in [43]), the energy function of a network equals the sum of the energies of its constituent nodes. This has significance to neural systems.

Another critical area is the consideration of the computational issue and how it can represent more complex problems and solution methodologies beyond those considered here. For example, an issue A can be logically manipulated just like any other question. To illustrate, if $A = A_1 \wedge A_2 \wedge \dots \wedge A_m$, then m parallel intelligent processes can be launched. One can consider how to partition a complex task into additional but simpler tasks.

This paper only discusses discrete memoryless intelligent processes. More complex processes can be constructed from this building block. For example, a complex task can be composed from simpler tasks, as is also illustrated by $A = A_1 \wedge A_2 \wedge \dots \wedge A_m$. As described in [1], biological systems as intelligent processes are differentiated from other types because of their hierarchical nature.

Lastly, one can consider varying intelligent process types. The cortical neuron is an example of a system that acquires information from its environment and then uses it to make decisions. One can simply have a question-asking process similar to that in a card-guessing game that is entirely information oriented. Many games are like this. Conversely, one can have a process that is purely self-constructing or control oriented, such as in constructing an organism from its DNA sequence.

6. Conclusions

This paper argues that basic physical and information-theoretic principles may underlie intelligent processes.

First, it was described that one can develop a basic theory of computation describing how information and control are exchanged by a system with its environment. This theory is based on computationally interpreting distinguishability and causality. Distinguishability leads to the consideration of a joint logic of questions and assertions as developed by Cox [5]. Causality leads to the conclusion that only two kinds of computational processes are possible; ones that reconstruct the past and ones that control the future. Intelligent systems fall into the latter class.

Next, the engineering mathematics of the min-max optimization process required to realize double-matching optimization by an intelligent process is summarized. Double-matching is thought to underlie the optimization strategy of intelligent systems. While optimizing its computational capacity, it also optimizes its energy efficiency.

Finally, this paper demonstrates the engineering process of reverse-engineering a cortical neuron using the notion of questions developed initially, and then applying the requirement of double-matching as its adaptation strategy. This led to a physiologically accurate prediction of known neural functionalities. These include signaling by action potentials, Hebbian adaptation, dendritic field, axonal decisioning, the refractory period, and many other known properties of cortical neurons.

Should the proposed model of intelligent processing be correct, then there are many implications. This paper has not even begun to consider many of the next steps to be taken and questions to be asked. While these questions are both scientific and technical, they are also philosophical.

Regarding philosophical implications, consider the question “If a tree falls in the forest and nobody is there to hear it, then does it make a sound?” The theory here says that if is not answered by any real assertion then it is known as a vain question. Therefore, this philosophical question has an answer and that answer is no.

Acknowledgments: The author wishes to thank Todd Hylton of Brain Corporation, who created and managed the Physical Intelligence program at DARPA and who proposed the *Thermodynamic Computing* program there. Thanks also to Alex Nugent, who likewise has contributed to the development of TCs through his creative and insightful nature.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are use in this paper:

GP	Geometric Programming
LP	Linear Programming
QSF	Quantum Synaptic Failure
TC	Thermodynamic Computer

References

1. Fry, R.L. Computation by Biological Systems. *Biomed. Sci. Today* **2015**. Available online: <https://pdfs.semanticscholar.org/c645/9f5948618a3773278514e67c05e02a35e9cf.pdf> (accessed on 6 March 2017).
2. Shannon, C.E. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949.
3. Online Encyclopedia of Science. Available online: <http://www.daviddarling.info/encyclopedia/P/protozoan.html> (accessed on 30 December 2016).
4. Lewis, C.I. *Mind and The World Order: An Outline of a Theory of Knowledge*; Charles Scribner’s Sons: New York, NY, USA, 1929; Reprinted in Paperback by Dover Publications, Inc.: New York, NY, USA, 1956.
5. Cox, R.T. Of inference and inquiry, an essay in inductive logic. In *Proceedings of the Maximum Entropy Formalism—First Maximized Entropy Workshop*, Boston, MA, USA, 1979; pp. 119–168.
6. Cohen, F.S. What is a question? *Monist* **1929**, *39*, 350–364. [CrossRef]
7. Lewis, C. *A Survey of Symbolic Logic*; University of California Press: Berkeley, CA, USA, 1918.
8. Cox, R.T. *Algebra of Probable Inference*; Johns Hopkins University Press: Baltimore, MD, USA, 1961.
9. Venn, J. On the Diagrammatic and Mechanical Representation of Propositions and Reasonings. *Philos. Mag. J. Sci.* **1880**, *10*, 1–18. [CrossRef]
10. Yeung, R.W. A new outlook on Shannon’s information measures. *IEEE Trans. Inf. Theory* **1991**, *37*, 466–474. [CrossRef]
11. Spencer-Brown, G. *Laws of Form*; Crown Publishing Group: Danvers, MA, USA, 1972.
12. The Shell Game. Available online: https://en.wikipedia.org/wiki/Shell_game (accessed on 6 March 2017).
13. Shannon, C.E. Source Coding with a Fidelity Criterion. *Proc. IRE* **1959**.
14. Cover, T.M. Untitled Contribution. *IEEE Inf. Theory Soc. Newslett.* **1998**, 18–19. Available online: <http://www-isl.stanford.edu/~cover/papers/paper112.pdf> (accessed on 6 March 2017).
15. Gibbs, W.J. A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces. In *The Scientific Papers of J. Willard Gibbs*; Longmans, Green, and Co.: New York, NY, USA, 1906; Volume 1, pp. 1–32.
16. Tribus, M. *Thermostatistics and Thermodynamics: An Introduction Energy, Information and States of Matter with Engineering Application*; D. Van Nostrand, Inc.: Princeton, NJ, USA, 1961.
17. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **1961**, *5*, 183–191. [CrossRef]
18. Bennett, C.H. The thermodynamics of computation—A review. *Int. J. Theor. Phys.* **1982**, *21*, 905–940. [CrossRef]

19. Gastpar, M.; Rimoldi, B.; Vetterli, M. To Code, or Not to Code: Lossy Source-Channel Coding Communication Revisited. *IEEE Trans. Inf. Theory* **2003**, *49*, 1147–1158. [CrossRef]
20. Chiang, M. Geometric Programming for Communication Systems. *Found. Trends Commu. Inf. Theory* **2005**, *2*, 1–154. [CrossRef]
21. Ben-Tal, A.; Nemirovski, A. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*; SIAM: Philadelphia, PA, USA, 2001.
22. Bertsekas, D.P. *Nonlinear Programming*, 2nd ed.; Athena Scientific: Belmont, MA, USA, 1999.
23. Bertsekas, D.P.; Nedic, E.; Ozdaglar, A. *Convex Analysis and Optimization*; Athena Scientific: Belmont, MA, USA, 2003.
24. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
25. Nesterov, Y.; Nemirovsky, A. *Interior Point Polynomial Algorithms in Convex Programming*; SIAM: Philadelphia, PA, USA, 1994.
26. Nocedal, J.; Wright, S.J. *Numerical Optimization*; Springer: New York, NY, USA, 1999.
27. Rockafellar, R.T. Lagrange Multipliers and Optimality. *SIAM Rev.* **1993**, *35*, 183–283. [CrossRef]
28. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.
29. Csiszar, I.; Korner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Academic Press: Salt Lake, UT, USA, 1981.
30. Phelps, R.R. *Convex Functions, Monotone Operators and Differentiability*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 1991; p. 42.
31. Avriel, M. (Ed.) *Advances in Geometric Programming*; Plenum Press: New York, NY, USA, 1980.
32. Wissner-Gross, A.D.; Freer, C.E. Causal Entropic Forces. *Phys. Rev. Lett.* **2013**, *110*. Available online: http://math.mit.edu/~freer/papers/PhysRevLett_110-168702.pdf (accessed on 6 March 2017).
33. Fry, R.L. Neural Statics and Dynamics. *Neurocomputing* **2005**, *65*, 455–462. [CrossRef]
34. Berger, T. *Rate Distortion Theory: A Mathematical Basis for Data Compression*; Prentice Hall: Upper Saddle River, NJ, USA, 1971.
35. The Nerve Cell. Available online: <https://www.britannica.com/science/nervous-system/The-nerve-cell#ref606326> (accessed on 6 March 2017).
36. Fry, R.L. Observer-Participant Models of Neural Processing. *IEEE Trans. Neural Netw.* **1995**, *6*, 918–928. [CrossRef] [PubMed]
37. Hebb, D.O. *The Organization of Behavior*; Wiley and Sons: New York, NY, USA, 1949.
38. Levy, W.B.; Baxter, R.A. Energy-Efficient Neuronal Computation via Quantal Synaptic Failures. *J. Neurosci.* **2002**, *22*, 4746–4755. [PubMed]
39. Hinton, G.E.; Sejnowski, T.J. Optimal Perceptual Inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, June 1983; pp. 448–453. Available online: <https://papers.cnl.salk.edu/PDFs/Optimal%20Perceptual%20Inference%201983-646.pdf> (accessed on 6 March 2017).
40. Oja, E. A simplified neuron model as a principal component analyzer. *J. Math. Biol.* **1985**, *15*, 267–273. [CrossRef]
41. Fry, R.L. Computation by neural and cortical systems. In Proceedings of the CNS Workshop on Methods of Information Theory and Neural Processing, BMC Neuroscience, Portland, OR, USA, 23–24 July 2008.
42. Fry, R.L. Double Matching: The Problem that Neurons Solve. *Neurocomputing* **2005**, *69*, 1086–1090. [CrossRef]
43. Hill, T.L. *Thermodynamics of Small Systems, Parts I & II*; Dover: New York, NY, USA, 2013.

