Noise, physics, and non-Turing computation

Sunny Bains[†] and Jeffrey Johnson
Department of Design and Innovation
The Open University
Milton Keynes, MK7 6AA, United Kingdom
[†]Corresponding author.
E-mail: sunny@sunnybains.com

Abstract
It has been proven that recurrent neural networks with analog weights can solve problems that are insoluble using Turing machines. However, the prevailing view of those involved in computing research is that a) non-Turing computation is unnecessary for intelligence and b) machines with non-Turing abilities cannot be built in practice. Two reasons are commonly cited for this. First, the effects of noise are said to remove any possibility of using an analog system to represent quantities with continuous infinite precision. Second, the (asserted) discrete nature of the Universe, is cited as proof that truly analog systems cannot exist. These are issues of physics that exist well outside the domain of computer science and engineering. Here, we will examine the technical and mathematical bases of these claims, suggest counter arguments, and discuss the deep physical questions that they raise.

Introduction
Siegelmann has shown that recurrent analog neural networks (RANN) [Siegelmann] have, at least theoretically, the ability to perform computational tasks that go beyond those of the Turing machine. This could help to justify an analog version of computationalism: the position that human-like intelligence is possible in computers. In particular, by removing computers from the confines of the Turing machine [Church] [Turing], many of Penrose's well-known objections to AI [Penrose] would become moot. In this sense, it is not a *counter* argument (of which there are many in the literature, see, for example [LaForte]), but rather a solution to the problem Penrose raises. ([Deutsch] suggests a partial solution based on quantum computing, but this construct—the universal quantum computer—is still not comprehensive in its power: it cannot compute non-recursive functions.)

The practicality of the super-Turing finding has been questioned. Some researchers claim that analog (as opposed to digital) computation is a meaningless concept because digital computers can always attain a sufficient level of precision to prevent there being any difference in outcome. Others claim that noise prevents real neural networks from reaching their alleged potential (see, for example, [Maass]). Though there are many such analyses, most have in common that they come directly from interpretations of Turing universality and communications theory.

This paper will discuss the validity of using such theories to describe continuously-evolving physical systems—whether electronic or biological—in which the processes taking place cannot be described in terms of conventional measurements. First, we will cover some basic arguments in favor of super- or non-Turing computation. Next, we will look at the objections to computation beyond the Turing limit. Then, we will examine the underlying physical assumptions that lead to these objections, after which we will describe what we believe to be a more physically valid interpretation of the way that recurrent analog neural networks interact with the outside world. Finally, we will explain how, as Penrose has always contended, the understanding and interpretation of quantum mechanics lies at the heart of the problem of computational intelligence: at least from a theoretical point of view.

The mathematical basis for super- and non-Turing computation
From a mathematical perspective, allowing neural networks to have weights that vary continuously clearly allows for a richer representational ability than a machine that uses a finite number of discrete symbols on an infinite piece of tape (part of the definition of a Turing machine). Though we will not go into the details here, the following mathematical arguments are straightforward to prove.

If a Turing machine is to process in a finite time, and each square or digit on the tape takes a finite time to read, then only a finite portion of the infinite tape can actually be used.[*] This limits the device to dealing with a subset of rational numbers (those rational numbers that can be represented by the particular number of symbols or bits that can be read in the allotted time). To represent a continuous-valued quantity, the set of real numbers, **R**, must be mapped on to (at most) the set of rational numbers **Q**. There are more irrational (of the set **Q'**) than rational numbers. Though both sets are infinite, the former has cardinality $\aleph_1$ whereas the latter has cardinality $\aleph_0$. As a result, many states in the

[*] *There is an argument that you can avoid this problem by using geometrically decreasing times, eg. processing the first digit in time **t**, the second in **t/2**, the third in **t/4**, etc. This produces a series that converges to **2t.** However, this has a problem of specificity: in this case opposite to that of the arbitrary precision argument discussed on the next page. In the former, the ability to process an infinite number of squares disappears as soon as one specifies a particular minimum processing time per square. In the latter the arbitrary precision is only sufficient as long as the data is specified in advance.*

original continuous function map to a single number within the Turing machine. This loss of information between the continuous-valued world and the Turing machine is the source of the proposed non-Turing abilities of the RANN (for a full proof and a discussion of what these abilities are, see [Siegelmann]).

Objections

For any given process, the Turing machine can be designed use enough bits to produce the correct answer. This is only true if the designer knows in advance what the required precision is. For nonlinear functions, the data itself determines how much precision is required. Take the sigmoid function, for instance, going from **–1** through **0** to **+ 1**. If the magnitude of the incoming continuous signal is relatively large (far from the threshold level **0**, then it can be represented very inaccurately without changing the output. For inputs of smaller magnitude, however, more precision is required. The difference between **+0.0001** and **-0.0001** in the input, may make a huge difference to the output. This problem increases as inputs get closer to zero. To correctly represent the incoming signal therefore requires foreknowledge of its value. While this is feasible for many engineering applications, it is not practical for intelligent systems designed to function in an unforeseeable universe.

Quantum mechanics says that the universe is discrete, not continuous, as evidenced by the existence of the Planck length and time. Thus, there is a finite level of precision beyond which it is meaningless to go. Even if length were discrete, position could still be a continuous variable. Imagine a watch with a second hand (of **n** Planck lengths, where **n** is an integer) tracing out a circle. Consider that a particle at the end of the second hand moves discretely: one Planck length in each Planck time. In order to arrive at its starting point (and take the exact same positions for the next cycle), the total distance traveled must be some integral number of Planck lengths. But it isn't: the circumference of a circle is $\mathbf{2\pi r}$, which in this case is $\mathbf{2\pi n}$ times the Planck length. Because π is an irrational number, not only there will there be an offset from the starting position in the first cycle to that of the second, but there will be an offset for every cycle. Further, the particle *cannot* go through the same position twice because it can only move an integral number of Planck lengths in distance. For it to take the same position on the $\mathbf{x^{th}}$ turn as it did on the first, therefore, $\mathbf{x2\pi n}$ has to be an integer. An irrational number multiplied by an integer is always irrational, so this is impossible. Thus, if the second hand is allowed to turn an infinite number of times it must pass through an infinite number of positions. Position must, therefore, be considered a continuous variable.

Regardless of the precision that a Turing machine can or cannot achieve, a practical analog machine can do no better because of noise. Where a digital computer loses information by mapping a continuous variable onto a discrete set of numbers, an analog computer loses information by mapping any given input point onto many output points (a different output for each noise value). The super- or non-Turing machine is therefore physically unrealizable. See, for example, [Hadley]. Though recent results [Mar] have shown that, in fact, coupled neural networks are much less sensitive to noise than had previously been thought, this objection gets to the heart of the controversy concerning super-Turing computation. One weakness of this position that no-one has, so far, proven that noise removes all *non*-Turing abilities from the RANN (though it has been proven that some of the Turing abilities are lost, eg. [Maass]). It is possible that we are not forced to choose simply between Turing and super-Turing abilities, but that there is a realizable subset of the latter that can perform *some* Turing and *some* non-Turing functions. However, this question remains unanswered, so it does not really help us here.

The analysis of noise as a degrading element in recurrent analog neural networks is a natural result of communications theory [Shannon]. Communications theory is an extremely important foundation for many fields in computer science and electrical and electronic engineering. It is an important tool, but relies on the following assumptions (see Figure 1A):

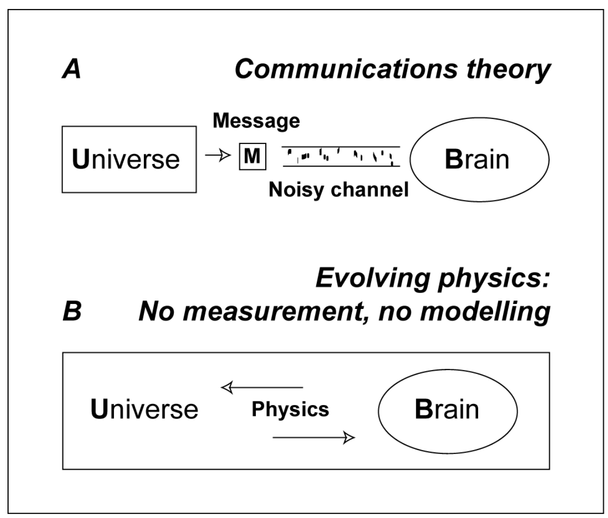

Figure 1: (A) The brain as a receiver and processor of information that comes from a sender (the Universe) via a noisy channel. (B) The brain and Universe as two coupled, co-evolving physical systems.

1. There is a *sender* who is trying to convey a particular piece of information or *signal* down a *channel* containing *noise* to a *receiver*.
2. The noise is unrelated to the sender or receiver.

Using communications theory to analyze a digital computer or analog communications system is appropriate. However, in the next section we will discuss why a recurrent analog neural network can be considered to be a qualitatively different kind of system or process: one that must be analyzed in a different way.

Physical computation

Consider a brain, whether electronic or biological, as having the following properties: it obeys the laws of physics and it is structured in such a way that it is particularly sensitive to, and able to adapt to, the rest of the universe. Consider that, at time *t* the brain is in state **B**. By this I mean that, were it possible to measure all the physical variables relating to all the particles that comprise electrons, atoms, molecules, neurons etc., respectively, they would be described by **B**. (The fact that this quantity is, in principle, unmeasurable, is irrelevant to our argument). Consider that, at time *t*, the rest of the universe could be described by a similar state variable **U**. These two "entities"—brain and universe—exert forces on each other and exchange particles and energy as dictated by the laws of physics (see Figure 1B).

Let us now assume that the universe is *deterministic* and *non-random* (we will come back to this assumption later). In this case, as **B** and **U** evolve over time, they will be intrinsically coupled. By this I mean that if "outside observers" had knowledge of **B** at time t_1, and complete understanding of the laws of physics, then by looking at **B** again at time t_2 they could infer something about **U** without observing anything but the brain. This information would be incomplete or ambiguous, in that many different states of the universe could cause the same change in the way the brain evolves with time. However, there is no sender, no message, no receiver, no channel, and therefore no channel noise here: simply two co-evolving physical systems.

The aforementioned ambiguity is the only source of information loss in this situation. Noise is not an issue at this point because it is a property that can only emerge through a *statistical model* of some phenomenon (such as thermodynamics). In such models, we make an engineering trade-off: we free ourselves of the need to know the *microscopic details* of a situation (like the position and momentum of individual gas molecules) in exchange for some *uncertainty* in our ability to predict. There are no such compromises made in our scenario. All variables are specified, all forces are known. All that the universe and brain have to do is to obey the laws of physics and they will evolve from their initial conditions with *infinite precision.*

At this point we should point out that the situation outlined here is as true for an inanimate object as it is for the brain. The object, left to its own devices, will evolve exactly as the laws of physics dictate. There is no opportunity for information to be lost.

Next, consider what makes the biological brain, for instance, different from an inanimate object: the fact that the former is *physically structured to evolve in such a way that its ability to survive increases over time*, allowing it (and the animal that's attached to it) the opportunity to procreate. *(This is not to suggest that intelligence is a passive process, merely that the activity is determined, like everything else, by physical laws)*. Let us consider, that the brain is structured as a recurrent analog neural network (RANN) and includes sensory organs (ie. that the brain is *embodied*). First, let us note that just because the system is designed to be particularly adaptive to particular sorts of stimuli (eg. light, sound), that does not mean that it is *only* sensitive to these phenomena. Gravity, invisible radiation (to the eye), magnetic fields etc. all have some influence on the state of the matter that makes up the brain: even if that influence is small (affecting microscopic rather than macroscopic events). Though the effect may be subtle, these events will affect the operation of the neural network. Because learning is going on in the presence of this underlying microscopic activity, that must somehow be reflected in what is learned. Thus these phenomena represent *a source of information about the universe* just as much as incoming light rays do.

Second, the brain *has no way of distinguishing between "signals" and anything else, except by experience.* The brain's job is merely to map the right set of responses to the external conditions, however those happen to manifest themselves. Information about temperature, for instance, encoded as what engineers might call "thermal noise", may be just as important for determining a particular outcome as whether the sky is light or dark. Thus, "thermal noise" is a *signal* for that task. (Likewise, there may be cases where, for instance, aural inputs are irrelevant to the task at hand and the brain can learn from experience to ignore them for that purpose). Thus what constitutes noise is purely context dependent in the physical brain.

In other words, it is only when we attempt to *model* these brain functions that we are forced to describe as noise those things that we *cannot* model: either because they are unmeasurable or because they complicate the calculation. Looking from the opposite point of view, consider the engineer trying to model a particular neural network. One can model the function of an individual neuron, the constituent molecules, the electrochemical components, down to the quantum-mechanical level. At each stage uncertainty (or noise) is stripped away, the trade-offs being increased complexity in the model and increased precision required of the initial conditions. At the lowest level, that of quantum mechanics (and related theories), the noise disappears entirely (as everything is explicitly known) except for quantum noise.

For the engineer, this precision is of course impossible. According to the uncertainty principle, there is no way to make the measurements required to predict the behavior of the neural network. However, the engineers only job is to make sure that the neural network (or brain) is sensitive to the universe at large and able to adapt in useful ways. After that, physics takes over. There are no explicit measurements, no calculations: the analog network can simply evolve according to the physical laws (which it must, and will, do precisely). It is this *physical* computation that is not subject to conventional types of noise.

The statistical nature of Quantum Mechanics

If the super- or non-Turing properties of recurrent analog neural networks are still to be eroded by the presence of noise, that noise must be quantum mechanical. Further, that quantum mechanical noise may not simply be caused by *observational unpredictability* (where the observer simply lacks enough information determine what will happen next) or problems of *measurement* (where the measurement itself changes the system, thus preventing the original quantity of interest being accurately determined). In order for the quantum noise to be truly noisy, it must be unpredictable by physics itself.

There are two interpretations of quantum mechanics that allow this true unpredictability. First, the so-called "many worlds" theory suggests that all statistical possibilities for any particular quantum event do in fact happen. Our universe simply takes a "random walk" through these events. Another possibility is that there is some intrinsic source of randomness in the universe: so that some kind of "coin toss" takes place to determine every quantum event. These true noise sources would act to "decouple" the evolving systems, both brain and universe, degrading the information that each had obtained from its physical interaction with the other.

However, there are also theories (see, for example, [Bohm]) that allow for quantum noise to be unpredictable to observers but not to physics. The uncertainty principle prevents us from making sufficient measurements to make good predictions but, according to these theories, the underlying physical properties are still there and determining the outcome. No "coin" must be flipped. physics merely evolves according to quantities we cannot measure and laws we don't yet understand. In this situation, there is no such thing as noise.

Discussion and further work

The physics community currently favors the "random" interpretation of quantum mechanics, which could indeed prevent the existence of super-Turing computation. However, quantum mechanics is still not well-understood as a physical phenomenon (as opposed to a mathematical framework) so it is premature to use this result to answer the computational question. In addition, there seem to be physical phenomena that perform functions (or computations) that are computable neither by Turing machines nor by universal quantum computers. Quantum computation over continuous variables has been proposed, but because these systems still involve programming (setting initial states) and then explicit measurements, they have been found to be equivalent to the universal quantum computer [Lloyd]. Further examination of the non-simulable should shed light on the prospects for physical (non-algorithmic) computation.

The authors would like to thank the many scientists and engineers who generously contributed to this work through reference, discussion, and criticism.

References

Bohm, D (1952), *A suggested interpretation of the quantum theory in terms of "hidden" variables,* **Phys. Rev. 85,** pp. 166-193.

Church, A (1936), *An unsolvable problem of elementary number theory,* **American Journal of Mathematics 58,** pp. 345-363.

Deutsch, D (1985), *Quantum theory, the Church-Turing principle and the universal quantum computer,* **Proc. R. Soc, Lond. A 400,** pp. 97-117.

Hadley, R.F (1999), *Cognition and the Computational Power of Connectionist Networks,* **Proc. 21st Ann. Conf. of the Cognitive Science Society,** Vancouver, Lawrence Erlbaum Assoc..

LaForte, G (1998), P J Hayes and K M Ford, *Why Gödel's Theorem Cannot Refute Computationalism,* **Artificial Intelligence 104** (1-2), pp. 265-286, September.

Lloyd, S and S L Braunstein (1999), *Quantum computation over continuous variables,* **Phys. Rev. Lett. 82** (8), 22 February.

Maass, W (1997), *Analog Neural Nets with Gaussian or other Common Noise Distributions cannot Recognize Arbitrary Regular Languages,* **Electronic Colloquium on Computational Complexity Report TR97-052.**

Mar, D J (1999), C C Chow, W Gerstner, R. W Adams, and J J Adams, *Noise shaping in populations of coupled model neurons,* **Proc. Natl. Acad. Sci. USA 96,** pp. 10450-10455, August.

Penrose, R (1989), ***The Emperor's New Mind,*** Oxford University Press.

Shannon, C E (1949), W Weaver, **The mathematical theory of communication,** University of Illinois Press.

Siegelmann, H T (1999), ***Neural Networks and Analog Computation: Beyond the Turing Limit,*** Birkhäuser Boston.

Turing, A M (1936), *On computable numbers, with an application to the Entscheidungsproblem,* **Proc. London Mathematical Society, Series 2, 42,** pp. 230-265.