TITLE

Consciousness as an emergent property of non-deterministic computation in cortical neural networks

AUTHOR

Elizabeth A. Stoll, Ph.D.

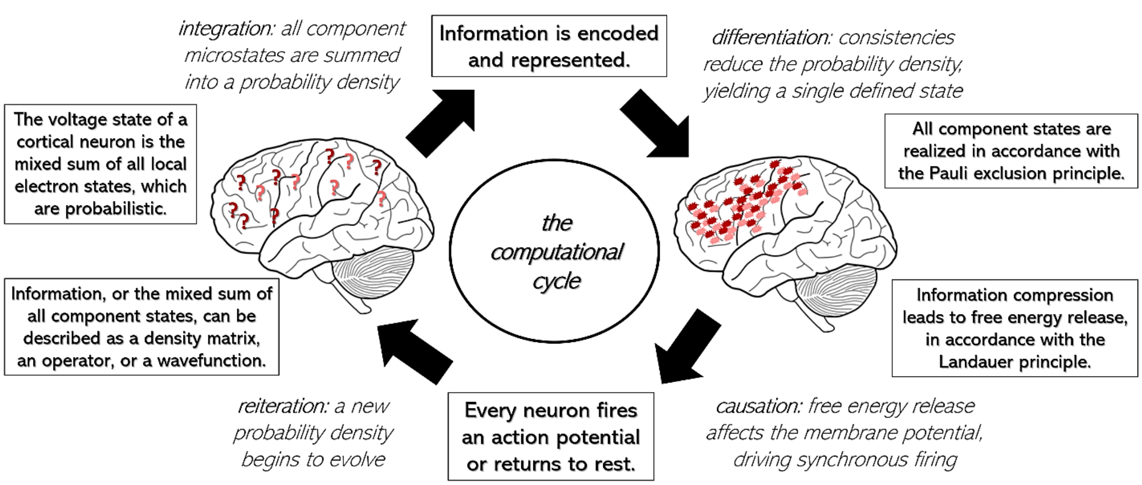
ABSTRACT

Problem: The field of neuroscience is currently unable to describe why cortical neural networks – but not spinal reflex circuits – are associated with perceivable information content, predictive models of the world which grow over time, and the initiation of voluntary behavior.

Methods: This project employs established mathematical toolkits in computational physics to model cortical neuron signaling outcomes. Typically, the cortical neuron is modeled as a binary computational unit, either spiking or not spiking (thereby encoding Shannon entropy). Here, the cortical neuron is modeled as a two-state quantum system, with some probability of switching from an off-state to an on-state (thereby encoding von Neumann entropy). This approach takes into account the contribution of both upstream signals and random electrical noise in gating signaling outcomes. With this method, the membrane potential is described as a macrostate – the mixed sum of component microstates, or the quantity of information that is physically encoded by the computational unit.

Results: Three independent mathematical models each demonstrate how cortical neurons engage in a cyclical process of information generation and compression, as free energy is distributed toward some thermodynamic quantity of information, then partially recovered as predictive value is extracted during this non-deterministic computation. The computational process results in synchronous firing across the network. Calculations for coulomb scattering profiles, decoherence timescales, and the energy efficiency of the brain are shown to be compatible with this process of thermodynamic computation.

Conclusions: This new approach to modeling probabilistic signaling outcomes in cortical neurons naturally yields representative information content, the spontaneous emergence of a more ordered system state during predictive processing, and contextually-appropriate behaviors resulting from a system-wide non-deterministic (but fully mechanistic) decision process. More generally, this approach yields a theoretical framework for semi-Markovian systems achieving Bayesian inference at any scale, through a hardware-instantiated non-deterministic computation that cannot be replicated in classical computing architecture. This approach has only three (empirically-valid) assumptions, provides superior explanatory power than previous efforts to model the emergent properties of cortical neural networks, and offers multiple testable hypotheses for laboratory researchers. This new theoretical framework is expected to have a useful impact on the field of neuroscience and the field of computational intelligence.



KEY CONCEPTS

Entropy is a formal mathematical description of the possible arrangements of particles within a thermodynamic system. It is also a thermodynamic quantity, given by the amount of *energy* which is unavailable to do *work* in that thermodynamic system.

Information is also a formal mathematical description of possible system states. The equations for information (derived by Shannon and von Neumann) are nearly identical to the equations for entropy (derived by Boltzmann and Gibbs). The relationship between these thermodynamic quantities and these computational quantities is still being explored.

In **classical computing architecture**, each computational unit (or 'bit') has a binary '0' or '1' state. The macrostate of the system is the sum of all these component microstates, adjusted for the statistical likelihood of occurrence. This distribution of possible macrostates is the amount of 'information' held by the system.

In **modern quantum computing architecture**, each computational unit (or 'qubit') has a spin state *with some probability* of being either '0' or '1'. The macrostate of the system is the sum of all these component pure states, adjusted for their probability of occurrence. This distribution of possible macrostates is the amount of 'information' held by the system.

In this newly-proposed model of **ambient-temperature quantum computing architecture**, each computational unit (or 'electron') *has some probabilistic position and momentum* in relation to each computational unit, and each computational unit (or 'neuron') *has some probability of switching* from an off-state to an on-state. Here, the state of each neuron in the network is dependent on both upstream signals and random electrical noise. The 'information' held by each neuron is encoded into its membrane potential, with the macrostate of that computational unit being the mixed sum of all component pure states.

'Component pure states' are the distribution of eigenvalues for each electron in the system. An electron is not defined in the present moment; it is better described as having some distribution of possible values for position and momentum, rather than an exact value. The state of an electron may contribute to the membrane potential of multiple neurons. As these **component pure states** are resolved, the state of each neuron in the system is resolved.

The result of this computational process is 'populational coding' – the synchronous but statistically random firing of neurons across the network, at regular intervals, corresponding with the realization of perceptual content from all available sensory modalities. **Populational coding** is observed in neural networks that retain sensitivity to random electrical noise (such as the mammalian neocortex), but not in neural networks that are robust to random electrical noise (such as spinal reflex circuits).

Populational coding is well-established to be a neural correlate of perceptual content. But the mechanisms underlying the synchronous but statistically random outcomes for neurons across the cerebral cortex has not been well-understood. This new theoretical framework provides a *mechanism* for information to be both encoded and represented by a neural network, through an **inherently probabilistic** process of non-deterministic computation.

As probability amplitudes constructively and destructively interfere, resolving the system state in the present moment, information is physically compressed and each computational unit has an inherently probabilistic signaling outcome. If the membrane of each computational unit also meets the criteria for **a holographic recording surface**, then any encoded information will be paired with a holographic projection of that information content. This content is exclusively accessed by the encoding structure and is representative of incoming sensory data in all available modalities.

PROJECT DESCRIPTION

Introduction

Consciousness is typified by a cohesive stream of rich, qualitative perceptual experience; the emergence of cognitive models of the world, centered on the self; and the ability to initiate volitional action. All three of these features are tied to neuronal activity in the cerebral cortex. The qualitative nature of perceptual experience is limited by the range and sensitivity of the available sensory apparatus, and dependent on the cortical encoding of sensory stimuli [1-3]. Experiences drive synaptic remodeling and the formation of memories [4, 5], molding perceptual expectations in relevant contexts [6, 7]. These memories and expectations are then used to select contextually-appropriate behavior [8].

A mechanistic approach to understanding consciousness is needed, focusing on the unique way that cortical neurons encode information. Neurons in spinal reflex circuits have purely deterministic firing patterns, with signaling outcomes that can be easily predicted by summing upstream inputs [9, 10]. Meanwhile, cortical neurons have unpredictable signaling outcomes, with partially-stochastic charge flux and spontaneous subthreshold fluctuations in membrane potential substantially contributing to the likelihood of firing an action potential [11-13]. Neurons across the cerebral cortex actively maintain a coordinated 'up-state', with each cell suspended near action potential threshold, allowing random electrical noise to affect the probability of switching from an off-state to an on-state [14].

Methods

Previous approaches to modeling cortical neuron signaling outcomes

To compute the most likely state of the surrounding environment, a cortical neural network must select an optimal system state in the present context, from a large probability distribution. Researchers have previously modeled this inherently probabilistic computation with Bayesian statistics [15], random-connection models [16], or fanofactor analysis of spike variance over time [17]. While these purely statistical methods can model the inherently probabilistic behavior of cortical neurons at the population level, they do not explain the *mechanistic process* underlying a state change in individual cells.

The Hodgkin-Huxley equations provide a good approximation for predicting cortical neuron firing patterns under steady-state conditions [18]. Repetitive firing patterns even emerge in this model when differentiating with respect to time and applied current density [19, 20]. Linear approximations of the Hodgkin-Huxley model also accurately predict shifts in membrane potential, as long as temperatures are below 27°C and the region of membrane being modeled is above 200 square microns in size [21]. But it is worth noting that the underlying relationship between membrane voltage, ion conductances, and channel activation – given by these four partial differential equations – must ultimately be described by either modeling all eigenvectors in the system along real and complex planes, or by modeling a Hopf bifurcation to find the critical point where the cortical neuron flips from an off-state to an on-state [22-24]. Both of these deeper models essentially describe quantum processes, utilizing imaginary axes to approximate the contribution of inherently random events. Indeed, any attempt to achieve more accurate models of probabilistic cortical neuron firing runs into quantum mechanics.

A new approach to modeling cortical neuron signaling outcomes

The new approach formalizes the contribution of random electrical noise to cortical neuron signaling outcomes. The voltage state of a cortical neuron changes over time, as upstream signals and inherently stochastic events converge to affect the membrane potential. So, although neurons are classically described as binary computing units, encoding Shannon entropy, a cortical neuron is better described as having some probability of switching from an off-state to an on-state after some time evolution, thereby encoding von Neumann entropy. Here, the membrane potential of a cortical neuron is modeled as the uncertainty in all component pure states, or the amount of information encoded by that computational unit. This model of quantum coding is well-established in the literature [25, 26], and is applied here to cortical neurons.

The neuron starts in off-state ϕ, not firing an action potential, and after time t, it is in another state χ. Over this period of time, the state of the neuron evolves from ϕ to χ, with the probability of switching state given by the Kronecker delta. Three different mathematical toolkits are then employed to model how the system state evolves over time. By modeling the contribution of inherently probabilistic events to the probability of firing an action potential, the present method offers a fully mechanistic approach that accounts for both random and non-random contributions to the membrane potential.

Results

A model of non-deterministic computation in terms of matrix mechanics

Probabilistic component pure states can be represented algebraically by a density matrix [27]. The density matrix then undergoes a unitary change of basis, as the system state is perturbed by its surrounding environment over some time evolution. The diagonalization of the density matrix yields a zero determinant, assigning eigenvalues for all component microstates at some defined point in time. This computational process of identifying linear correlations in a combined density matrix is shown to be equivalent to the extraction of predictive value from a thermodynamic quantity of information. The system state transitions from a prior probability (the 'past') to a posterior probability (the 'present') through a time-dependent unitary change of basis, thereby achieving Bayesian inference. This model of matrix mechanics produces testable predictions regarding the coulomb scattering profiles and decoherence timescales of sodium ions at the neuronal membrane [27].

A model of non-deterministic computation in terms of Hamiltonian mechanics

Component pure states are integrated into a physical quantity of information, by populating a Hamiltonian operator [28]. The Hamiltonian is then differentiated with respect to all perturbations to the system, yielding eigenvalues, or observables, on the boundary region of that high-dimensional probability density. In accordance with the Hellman-Feynman theorem, the resolution of the system state is paired with a spontaneous shift in charge distribution, as energy is redistributed across the system. This newly-actualized system state immediately becomes the past, as a new probability density emerges in the present moment. This model of Hamiltonian mechanics produces testable predictions regarding the wavelength of free energy released upon information compression [28].

A model of non-deterministic computation in terms of wave mechanics

Component pure states can be represented geometrically as complex-valued waves or probability amplitudes [29]. Here, each electron is modeled as an electromagnetic point source, interacting in a probabilistic manner with each neuronal membrane. The electron is described as some set of probability amplitudes, distributed across five orthogonal axes: *x, y, z, atomic orbital,* and *time*. The membrane potential of each neuron is then a function of the probabilistic spatial position and atomic orbital of each local electron, after some time evolution – or the von Neumann entropy encoded by that computational unit. If the probabilistic trajectory of each electron over time t affects the voltage state of multiple computational units, then the system macrostate must be computed as a whole, with the state of every neuron being resolved as every component pure state is resolved. The constructive and destructive interference of high-dimensional probability amplitudes results in a non-deterministic computational outcome for each neuron, yielding an observable system state. If the surface of each computational unit is also a charge-detecting polymer substrate which meets the established criteria for a holographic recording surface, and it has the dimensionality to capture electrons as complex waves, then this encoding process will naturally generate a high-dimensional holographic projection of representative information content. The information content is exclusively accessed by the encoding structure, and representative of incoming sensory data, with the content limited by the range and sensitivity of the sensory apparatus. This model of wave mechanics produces testable predictions regarding the effects of various drug classes, electromagnetic stimulation, and alterations of lipid membrane composition on perceptual content [29].

Obeying the laws of thermodynamics

Critically, both upstream inputs and inherently probabilistic events contribute to the voltage state of a cortical neuron and the probability of it firing an action potential. The present methodology makes the following logical assertion: that macro-scale computational units which 1) expend free energy to physically encode information into the voltage state of a computational unit, and 2) allow noisy events to affect the probability of a signaling outcome, <u>must physically generate information entropy</u>. That physical quantity of information is parsed for consistencies and an optimal system state is actualized in the present context, thereby encoding the state of the local environment. The physical compression of information releases free energy, locally to any reduction of uncertainty in component states, in accordance with the Landauer principle [30-33]. Due to the relationship between the Gibbs free energy equation and the Nernst equation, the release of thermal free energy upon information compression directly affects signaling outcomes. Neurons which have reduced uncertainty locally <u>gain</u> free energy, allowing them to restore their resting potential. Neurons which have gained uncertainty over the course of the computation <u>lose</u> free energy, and move closer to action potential threshold. For cortical neurons in up-state, these fluctuations dominate membrane potential dynamics and drive signaling outcomes.

This system-wide computation results in sparse but synchronous firing across the entire neural network, paired with the realization of a percept. Over time, as the system extracts predictive value from a thermodynamic quantity of information, it spontaneously remodels itself into a more ordered system state to store this predictive value. This computational process thermodynamically favors that same pattern of activity to re-occur in a similar context or with a familiar stimulus - prompting an organism to explore its environment, thereby generating information with potential predictive value, and to exploit any knowledge gained, by utilizing previously-acquired predictive value in a familiar context.

The net amount of energy entering the system provides a strict thermodynamic limit on the amount of information that can be generated *and* the amount of predictive value that can be extracted over t [34]. The extraordinary energy efficiency of the human brain is consistent with this process of thermodynamic computation, producing far too little entropy to obey classical assumptions [35].

The neural correlates of consciousness in this theory

1) If a far-from-equilibrium thermodynamic system traps heat to do work, this thermoregulated system may be able to take on a more ordered system state over time, by identifying linear correlations with its surrounding environment.

2) If a thermoregulated network of macro-scale computational units engage in probabilistic coding to physically encode information, allowing random noise to gate a signaling outcome, this neural network will operate as a non-deterministic computer, solving problems through predictive processing.

3) If the outer membrane of the macro-scale computational units within that system *also* meet the criteria for a holographic recording surface, then encoded information will also be perceivable to the system encoding it. The information content that is experienced is a predictive statement about reality.

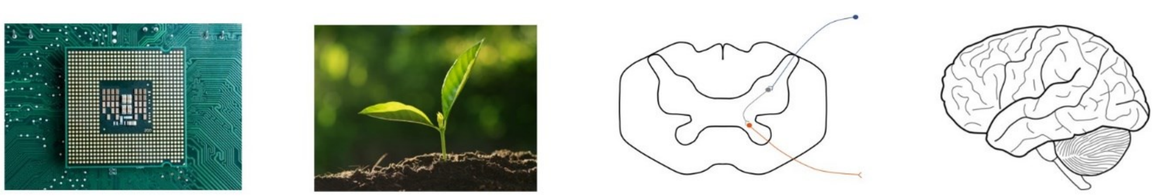
Any system meeting these three specific criteria should spontaneously initiate contextually-relevant yet non-deterministic behavior and perceive the information content being encoded.

Conclusions

This proposed process of thermodynamic computation allows an organism to effectively perceive and navigate its local environment, achieving exascale computational power with extraordinary energy efficiency. In this model, the key features of consciousness – including representative information content, predictive models, and non-deterministic yet contextually-relevant behavioral outcomes – emerge naturally from the activity of cortical neural networks.

Thermal noise reduces accuracy and energy efficiency during the computation.	Thermal noise must be kept out because it introduces errors in the computation.	Thermal noise is acquired from the local environment to fuel the computation.
Digital Bits: Each computational unit is either in 0 or 1 state	Qubits: Each electron has some probable 'up' or 'down' spin state	Qubits: Each electron has some probable location relative to each neuron
classical computation	SU2 quantum computing	SU4 quantum computing

Figure 2. A comparison of classical computation, modern quantum computation, and the process of ambient-temperature quantum computation that is proposed to take place in the cerebral cortex.



NO probabilistic coding to gate signaling outcomes	NO probabilistic coding to gate signaling outcomes	NO probabilistic coding to gate signaling outcomes	HAS probabilistic coding to gate signaling outcomes
NO holographic surface to encode information	NO holographic surface to encode information	HAS holographic surface to encode information	HAS holographic surface to encode information

Figure 3. A comparison of various thermodynamic systems, including a classical computer chip, a plant, a spinal reflex circuit, and the mammalian brain. Only the mammalian brain meets both key anatomical and physiological criteria for perceptual experience. Additional features required for conscious perception include: sensory input, motor output, and a thermoregulated system that traps heat energy to drive computational work. Furthermore, the development of a self-concept from an accumulation of lived experience requires extended periods of time to integrate information.

Appendix: Specific and testable predictions of the proposed theoretical framework

The theoretical framework produces the following testable hypotheses for laboratory neuroscientists:

1. **Coulomb scattering rate:** Coulomb scattering of sodium ions should be $\sigma \approx 32.8$ nm^2 at the cortical neural membrane. This prediction can be tested using cryo-electron microscopy.
2. **Decoherence timescales:** Decoherence timescales of sodium ions should be $\tau \approx 0.4$ ms at the cortical neural membrane. This prediction can be tested by measuring the leaky integration time, prior to firing.
3. **Neurophysiological correlates of information processing:** Spontaneous release of infrared photons should occur, with wavelengths of $\lambda \approx 46$ microns. This thermal free energy release should correspond to event-related potentials during perceptual tasks, but not ictal activity during seizures.
4. **Synchronous firing:** Synchronous firing should occur at nested frequencies between 1-200 Hz only in neural networks that engage in probabilistic coding and that trap heat to do computational work (non-dissipative, ergodic systems). This bursting neuronal activity should be abolished by absorbing infrared photon release, local to any reduction in uncertainty, during perceptual tasks.
5. **Spontaneous remodeling:** Spontaneous remodeling of the neural network into a more ordered state should occur only in neural networks that engage in probabilistic coding and that trap heat to do computational work (non-dissipative, ergodic systems). This remodeling activity should be abolished by absorbing infrared photon release, local to any reduction in uncertainty, during learning tasks.
6. **Extreme energy efficiency:** Systems that extract predictive value from a thermodynamic quantity of information should be far more energy efficient than any system operating under classical assumptions. Mammalian cortex, avian pallium, and any engineered system operating on the same principles should exhibit near 100% energy efficiency, while achieving unprogrammed exploratory behavior.
7. **Neuroanatomical correlates of information processing:** Individual subjective reporting of perceptual richness, on a customized scale, should correspond to different concentrations of polyunsaturated fatty acid and/or cholesterol content within the membrane of the computational unit. Altering the polyunsaturated fatty acid or cholesterol content of the cellular membrane, by modifying the dietary intake or the enzymatic machinery available in target cells, should alter perceptual content, leading to a corresponding increase or decrease in performance on stimulus-evoked tasks.
8. **Temperature effects:** Increased system temperature should *increase* perceptual content, leading to faster responses, higher error rates, and fatigue. Conversely, reduced system temperature or metabolic inefficiencies should *reduce* perceptual content, leading to impairments in neural network remodeling and initiation of volitional action, since free energy is dissipated before it can be directed toward work.
9. **Pharmacological effects:** Increasing excitatory currents should render cortical neuron signaling outcomes *more deterministic*, rendering cells *more likely* to fire. This experimental manipulation, using NMDA or AMPA receptor agonists, should lead to low-quality but high-intensity perception, reduced memory formation, and motor rigidity. Increasing inhibitory currents should also render cortical neuron signaling outcomes *more deterministic,* rendering cells *less likely* to fire. This experimental manipulation, using drugs that potentiate GABA receptor activation such as barbiturates or benzodiazepines, should lead to low-quality and low-intensity perception, reduced memory formation, and motor slackness. In both cases, restoring the probabilistic nature of cortical neuron activity should reverse these pharmacological effects, thereby restoring perceptual content, memory storage, and voluntary behavior. By contrast, any drug that increases subjective perceptual content, such as psilocybin, should do so via a mechanism of increasing the uncertainty of any signaling outcome, indicated by increased EPSCs and IPSCs in the 0.4 millisecond time window prior to firing an action potential.
10. **Electromagnetic effects:** Electrical or magnetic stimulation of neurons should have no discernable effect on perceptual content, if the field is uniform; stimulation should have some discernable effect if it is directed at a small population of computational units, altering the probability of them firing.

References

1. Pins, D., & Ffytche, D. (2003). *Cereb Cortex* 13(5): 461-74.
2. Kerlin, J. R., et al. (2010). *J Neuroscience* 30(2): 620-8.
3. Malnic, B., et al. (1999). *Cell* 5(5): 713- 23.
4. Hebb, D. O. (1953). *Br J Anim Behav* 1:43-7.
5. Zarnadze, S., et. al. (2016). *Elife* 5:14912.
6. Melloni, L., et al. (2011). *J Neuroscience* 31(4): 1386-96.
7. Kok, P., et. al. (2013). *J Neuroscience* 33(41): 16275-84.
8. Khalighinejada, N., et. al. (2018). *Neuroimage* 165: 35-47.
9. Bialek, W. & Rieke, F. (1992) *Trends Neurosci* 15(11): 428-34.
10. Powers, R. & Binder, M. (1995) *J Neurophysiology* 74(2): 793-801.
11. Stern, E.A., et. al. (1997) *J Neurophysiology* 77(4): 1697-715.
12. Dorval, A.D. & White, J.A. (2005) *J Neuroscience* 25(43): 10025-8.
13. Insanally, M.N. et al. (2019) *Elife* 8: 42409.
14. Steriade, M., et. al. (2001) *J Neurophysiology* 85(5): 1969-85.
15. Beck, J.M., et al. (2008) *Neuron* 60(6): 1142-52.
16. Maoz, O., et al. (2020) *Proc Natl Acad Sci* 117(40): 25066-73.
17. Fayaz, S., et al. (2022) *PLOS Comput Biol* 18(7):e1010256.
18. Hodgkin, A.L. & Huxley, A.F. (1952). *J Physiology* 117(4): 500-44.
19. DeFelice, L.J. & Isaac, A. (1992). *J Statistical Physics* 70: 339-54.
20. Pankratova, E.V., et. al. (2005). *European Phys Journal B* 45: 391-7.
21. Steinmetz, P.N., et al. (2000). *J Comput Neurosci* 9(2): 133-48.
22. Rinzel, J. & Miller, R.N. (1980). *Math Biosci* 49: 27-59.
23. Rowat, P. (2007). *Neural Comput* 19(5): 1215-50.
24. Austin, T.D. (2008). *Ann Appl Prob* 18: 1279-1325.
25. Schumacher, B. (1995) *Phys Rev A* 51: 2738.
26. Bennett, C.H. & Shor, P.W. (1998) *IEEE Trans Inf Theor* 44(6): 2724-42.
27. Stoll, E.A. Pre-print. doi:10.1101/2022.12.03.518981
28. Stoll, E.A. Pre-print. doi:10.1101/2022.12.03.518983
29. Stoll, E.A. Pre-print. doi:10.1101/2022.12.03.518989
30. Landauer, R. (1961) *IBM J Res Develop* 5: 183-91.
31. Berut, A., et al. (2012) *Nature* 483(7388): 187-9.
32. Jun, Y., et al. (2014) *Phys Rev Lett* 113: 190601.
33. Yan, L.L. et al. (2018) *Phys Rev Lett* 120: 210601.
34. Stoll, E.A. Pre-print.
35. Stoll, E.A. Pre-print. doi:10.1101/2022.12.03.518978