

A universal tradeoff between energy, speed and precision in neural communication

Subhaneil Lahiri, Jascha Sohl-Dickstein and Surya Ganguli

The brain performs myriad computations rapidly and precisely while only consuming energy at a rate of 20 watts, as opposed to typical supercomputers whose consumption rate is in the megawatt range. This remarkable performance suggests that evolution may have simultaneously optimized energy, speed and accuracy in neural computation. What are the fundamental limits and tradeoffs involved in simultaneously optimizing these three quantities, and how close do neural systems come to these limits? Information theory and thermodynamics provide upper limits on the accuracy and energy efficiency of any physical device, including biological systems. However, these bounds ignore the time it takes to achieve such limits and are therefore not obviously relevant to neurobiology, as neural systems cannot wait an infinite time to receive and transmit information.

We go beyond classical information theory and thermodynamics to prove a new theorem revealing that the communication of an external signal via any signaling system is subject to a universal tradeoff between (1) the rate of energy consumption by the system, (2) the speed at which the signal varies, and (3) the precision with which the signal is communicated. This applies to arbitrary systems, for example synaptic spines transducing incoming neurotransmitters, retinal opsins transducing light, or cochlear hair cells transducing sound. We find simply that the product of speed and precision is upper bounded by energy consumption. Intuitively, this three-way tradeoff arises because any increase in speed at fixed precision requires the signaling system’s physical states to change rapidly, leading to increased energy consumption. Similarly any increase in precision at fixed speed requires a larger change in the probability distribution over signaling system’s states, which leads to increased energy consumption. Our newly discovered three-way tradeoff motivates new experiments to assess exactly how close neural systems come to simultaneously optimizing energy, speed and accuracy.

Additional Detail

Methods: The efficient coding hypothesis states that sensory systems maximize their mutual information with the input at fixed energy consumption. However, the use of Information Theory in this context is problematic, given that it typically works in the limit of large block lengths, essentially assuming a very large number of independent observations, whereas neurons only have a short time to respond to stimuli. Similarly, the bounds on energy efficiency coming from thermodynamics are saturated when the systems under consideration operate in a slow quasistatic limit. As neural systems are required to operate in finite time, far from the idealized quasistatic regime, these bounds are not relevant to them.

There is a dire need for work on the effect of finite time constraints on these systems. Whilst there has been separate work on finite time in Information Theory (e.g. [Polyanskiy et al. 2010]) and Thermodynamics (e.g. [Schmiedl and Seifert, 2007]), the connection between the two is still lacking. In the biophysics literature there has been work on two-way tradeoffs between energy and precision (e.g. [Endres and Wingreen, 2009]) or speed and accuracy (e.g. [Murugan et al. 2012]) in areas such as cellular chemosensation and kinetic proofreading, usually focused on specific models. For sensory adaptation, the three-way tradeoff has been studied for a very specific biochemical model [Lan et al. 2012]. **We provide the first example of a *universal three-way tradeoff governing any system*.** By simultaneously considering energy, speed and precision, we achieve greater relevance to neurobiology.

To maintain complete generality of our theory, we model the signaling substrate as an arbitrary Markov process, where the external signal couples to it by modifying the control parameters that determine the transition rates of the Markov process, as described in Figure 1. The external signal can then be reconstructed downstream by observing the state of this system. This formalism is general

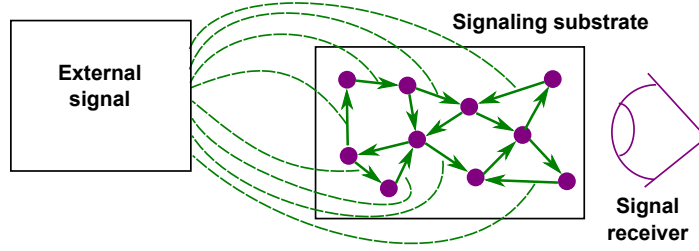


Figure 1: The physical substrate for signal transmission (e.g. a synaptic receptor complex or a molecular signalling cascade) is modeled as a stochastic dynamical system with transitions (green arrows) between its states (purple circles). The external signal (e.g. upstream neurotransmitter release or light) interacts with the system by modifying the transition rates, thereby influencing changes in the system’s state. Downstream processing can then estimate the external signal from the state of the signaling system. The precision of this estimate is limited by the energy consumption of the system and the speed of the external signal.

enough to incorporate many systems of interest in neuroscience, for example chemical networks underlying signaling cascades, or conformational changes of receptors and ion channels. In the example of a chemical network, the different signaling states would be different chemical concentrations and the external signal could be an enzyme or phosphorylation event that modifies the rate of conversion between different chemical species. In the example of receptors, the different conformational states would represent different signaling states, while the external signal could correspond to voltage changes or ligand binding events that modify the rate of various conformational changes.

Recent work in nonequilibrium statistical mechanics [Sivak and Crooks, 2012] uses linear response theory to express the energy consumption in terms of the velocity through control parameter space and a Riemannian metric tensor on that space. We derive a new relation between this metric tensor and the Fisher information of the system. We then further couple the Fisher information to estimation error, via the Cramer-Rao bound, thereby proving a new upper bound on the product of the squared speed of the signal and the precision of any estimate of the signal in terms of the energy dissipation rate:

$$[\tau \dot{\lambda}]^2 [\text{Var}(\hat{\lambda})]^{-1} \leq \frac{\tau P}{k_B T},$$

where τ is an intrinsic time constant of the signaling system, $\dot{\lambda}$ is the velocity at which the signal changes and $\text{Var}(\hat{\lambda})$ is the squared uncertainty of an estimate of the signal, $\lambda(t)$, from the state of the signaling system, P is the power, or rate of energy consumption, of the signaling system, T is the operating temperature and k_B is Boltzmann’s constant.

Intuitively to understand this result, we can think of $\mathcal{S} = [\tau \dot{\lambda}]^2$ as a measure of the external signal speed; it is the squared magnitude of the amount by which the external signal changes in one time constant of the signaling system. We can think of $\mathcal{P} = [\text{Var}(\hat{\lambda})]^{-1}$ as a measure of signaling precision; it is the inverse squared uncertainty of the signal estimate. And we can think of $\mathcal{E} = \frac{\tau P}{k_B T}$ as a dimensionless measure of energy consumption; it is the energy consumed in one time constant, relative to the energy scale set by the temperature. Then we have the simple non-dimensionalized inequality:

$$\mathcal{S} \mathcal{P} \leq \mathcal{E}.$$

All physical signaling systems, including all possible neural systems, are subject to this bound, and must therefore face a tradeoff between energy, speed and precision. This opens up the new experimental question of how energetically efficient, fast and precise neural communication is, relative to this bound.