Connectomic Constraints on Computation in Networks of Spiking Neurons

Venkatakrishnan Ramaswamy[1] *and* Arunava Banerjee[2]

Several efforts are currently underway to decipher the connectome of a variety of organisms. Ascertaining the physiological properties of all the neurons in these connectomes, however, is out of the scope of such projects. It is therefore unclear to what extent knowledge of the connectome alone will advance our understanding of computation occurring in these neural circuits. We consider the question of how, if at all, the wiring diagram of neurons imposes constraints on computation when we cannot assume detailed information on the physiological response properties of the neurons. We call such constraints, that arise by virtue of the connectome, *connectomic constraints* on computation. For feedforward networks equipped with neurons that obey a deterministic spiking neuron model which satisfies a small number of properties, we ask how connectomic constraints restrict the computations they might be able to perform. One of our results shows that all networks whose architectures share a certain graph-theoretic property also share in their inability in effecting a particular class of computations. This suggests that connectomic constraints are crucial for computation; merely having a network with a large number of neurons may not endow it with the ability to effect a desired computation. In contrast, we have also proved that with the limited set of properties assumed for our single neurons, there are limits to the constraints imposed by network structure. More precisely, for certain classes of architectures, we *must* have more detailed information on single neuron properties, before we can prove that there exist computations that the networks cannot perform. Thus, our theory suggests that while connectomic constraints restrict the computational ability of certain classes of network architectures, we may require more information on the properties of neurons in the network, before we can prove such results for other classes of networks.

Additional details. We begin by setting up a framework that allows us to view computation in feedforward networks as transformations that map input spike-trains to output spike-trains. We then ask what spike-train to spike-train transformations *all* feedforward networks of specific architectures cannot accomplish. In order to then rule out the prospect that the transformation in question is so "hard" that no network (of any architecture) can do it, we stipulate the need to demonstrate that there exists a network (of a different architecture) comprising simple neurons that can indeed effect this transformation. The goal of this work is to establish results of this form, after setting up a mathematical framework in which these questions can be precisely posed.

Our neurons are abstract mathematical objects that satisfy a small number of axioms that correspond to properties of neurons. Thus, our results are applicable to networks that use a neuron model that satisfies the same properties as well. The axioms used here simply assume that (a) Membrane potential is a function of input spikes received and output spikes emitted in a bounded past "window", (b) The neuron settles to the resting potential upon receiving no input spikes in the input window and having no past output spikes in the output window, (c) The neuron fires an action potential when its membrane potential hits, from below, a threshold (which may itself be a function of input spikes received and output spikes emitted in the bounded past window. (d) The neuron has an absolute refractory period, (e) Relative refractory period effects, i.e. presence of spikes in the past output window has a hyperpolarizing effect vis-à-vis absence of spikes in the output window, when spikes in the input window remain the same. Notably, no specific functional form of the membrane potential is assumed. Several widely-used neuron models such as the Leaky Integrate-and-Fire Model and the Spike Response Model satisfy the above properties, up to arbitrary precision.

We find that with the above properties, a single neuron cannot, in general, be rigorously treated as a transformation, which is to say a function, that maps input spike trains to output spike trains. Notwithstanding this, we prove that, under conditions consistent with spiking regimes observed in-vivo, the notion of a single neuron as a transformation is indeed well-defined and corresponds to maps from

[1]Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
[2]Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA.

finite-length input spike-trains to finite-length output spike-trains. However, these conditions turn out to be mathematically unwieldy when we try to tackle the question of what transformations networks with specific architectures cannot do. This motivates the development of a more tractable criterion; notably, we prove that this more tractable criterion can be used to prove every such result that is accessible to our more biologically well-motivated criterion, and therefore there is no loss of generality in using it.

Armed with this framework, we then proceed to set up definitions that allow us to ask what transformations networks of specific architectures *cannot* effect, that other networks can. Some results of this form are established; First, we show a transformation that a single neuron cannot effect but a network consisting of two neurons can. Next, we prove a result which shows that a class of architectures that share a certain abstract graph-theoretic property also share in their inability in effecting a particular class of transformations. Notably, while this class of architectures has networks with arbitrarily many neurons, we show a class of networks with just two neurons which can effect this class of transformations. This suggests that the network's structure is crucial in determining its computational ability.

While attempting to ask how increase in depth of the network constrains the transformations it can effect, we discovered that the current abstract model of the neuron does not adequately constrain networks in this respect. In particular, we prove that, with the current abstract model, every feedforward network, of arbitrary depth, has an equivalent feedforward network of depth equal to two that effects *exactly* the same transformation. The implication of this result is that we need to add more axioms to the present abstract model of the neuron in order for such results to be manifested. What this suggests is that the connectome alone may not impose strong constraints on the computational ability of certain types of networks. For such classes of networks, we might require more detailed knowledge about the properties of the individual neurons of the network, before we can characterize how the network's structure restricts its computational ability.

There has been considerable controversy about how useful data from the connectome projects might be in advancing our understanding of computation occurring in the circuits of the brain. While there is more to neuronal networks than just their wiring diagram, what our theory suggests is that the wiring diagram may impose crucial constraints on the computational ability of networks, in some cases. On the other hand, there also appear to be classes of networks for which a more detailed knowledge of single neuron properties might be necessary, before we can determine restrictions on their computational ability. With improving resolution of electron microscopy, the prospect of imaging single ion channels appears within reach. If successful, knowledge of the type and density of ion channels in each of the neurons could provide useful information that might help us tease out constraints on the computational capabilities of these networks.