J Comput Neurosci manuscript No.
(will be inserted by the editor)

Connectomic Constraints on Computation in Feedforward Networks of Spiking Neurons

Venkatakrishnan Ramaswamy · Arunava Banerjee

Received: date / Accepted: date

Abstract Several efforts are currently underway to decipher the connectome or parts thereof in a variety of organisms. Ascertaining the detailed 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 a mechanistic understanding of computation occurring in these neural circuits, especially when the high-level function of the said circuit is unknown. We consider, here, the question of how the wiring diagram of neurons imposes constraints on what neural circuits can compute, 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 if just by knowing the architecture of a network, we can rule out computations that it could be doing, no matter what response properties each of its neurons may have. We show results of this form, for certain classes of network architectures. On the other hand, we also prove that with the limited set of properties assumed for our model neurons, there are fundamental limits to the constraints imposed by network structure. Thus, our theory suggests that while connectomic constraints might restrict the computational ability of certain classes of network architectures, we may require more elaborate information on the properties of neurons in the network, before we can discern such results for other classes of networks.

Keywords Spiking neurons · Connectomics · Feedforward networks

1 Introduction

Recent remarkable experimental advances (Denk and Horstmann, 2004; Hayworth et al, 2006; Knott et al, 2008; Mishchenko et al, 2010; Turaga et al, 2010; Helmstaedter et al, 2011; Mikula et al, 2012) have brought the prospect of ascertaining the connectome or parts thereof closer to reality (Chklovskii et al, 2010; Kleinfeld et al, 2011; Seung, 2011; Denk et al, 2012; Reid, 2012; Helmstaedter et al, 2013). This data is currently not expected to include information on the detailed physiological properties of all the neurons in the connectome. Even so, already, there have been two pioneering studies (Briggman et al, 2011; Bock et al, 2011) that fruitfully use electron-microscopy reconstructions in conjunction with two-photon calcium imaging on the same tissue. In (Briggman et al, 2011), the authors used this approach to rule out certain models of direction selectivity in the retina. The other study (Bock et al, 2011) examined the orientation-selectivity circuitry in the cortex and found that inhibitory interneurons received convergent anatomical input from nearby excitatory neurons that had a broad range of preferred orientations. Recent work (Takemura et al, 2013) has also used connectomic reconstructions of the motion detection circuit in the fruit fly visual system, in order to identify cellular targets for future functional investigations; this is towards the goal of a comprehensive mechanistic understanding of this circuit. While this broad approach of combining functional imaging with structural reconstructions creates new opportunities to unravel structure-function relationships (Seung, 2011), to fruitfully

Computer and Information Science and Engineering,
University of Florida, Gainesville, FL 32611, USA.
E-mail: {vr1, arunava}@cise.ufl.edu

Present address of V. Ramaswamy:
Interdisciplinary Center for Neural Computation,
The Edmond and Lily Safra Center for Brain Sciences,
The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
E-mail: venkat.ramaswamy@mail.huji.ac.il

use functional imaging seems to require that (a) we have an a priori credible hypothesis about at least one high-level computation that the neural circuit in question is performing and (b) we have a way of experimentally eliciting performance of the said computation, usually via an appropriate stimulus. Unfortunately, neither of these conditions appear to be satisfied for a majority of neuronal circuits in the brain, especially as one moves away from the sensory/motor periphery. Suppose, in addition to its wiring diagram, we knew the detailed physiological response properties of all the neurons in such a neural circuit to the extent that we could predict circuit behavior (via simulations, for example). This might provide a way forward towards advancing hypotheses about what high-level computation(s) the circuit is actually involved in. Regrettably, ascertaining the detailed physiological response properties of all the neurons in such a network appears to be out of reach of current experimental technology. The prospects of obtaining the wiring diagram, however, seem to hold more promise. The question therefore becomes: (1) What can we learn from the wiring diagram alone, even when the specific high-level function of the neural circuit may be unknown? (2) Are there fundamental limits to what can be learned from the wiring diagram alone, in the absence of more detailed physiological information?

To investigate these questions, we have studied a network model equipped with neurons that obey a deterministic spiking neuron model. We ask what computations networks of specific architectures *cannot* perform, no matter what response properties each of their neurons may have. The implication, then, is that, owing to its structure, the network is unable to effect the computation in question. That is, connectomic constraints forbid the network from performing the said computation. In addition, to rule out the possibility that this computation is so “hard” that no network (of any architecture) can accomplish it, we stipulate the need to demonstrate that there exists a network (of a different architecture) comprising simple neurons that can indeed effect this computation. The goal of this paper is to establish results of this form for various network architectures, after setting up a mathematical framework within which these questions can be precisely posed. As a first simplifying step, in this paper, we limit our study to feedforward networks of neurons. Having started with this goal, however, we also find that with the small number of basic properties assumed for our model neurons, there are fundamental limits to the computational constraints imposed by network structure, in certain cases. In particular, we prove that, constrained only by the properties in the current neuron model, every feedforward network, of arbitrary size and depth, has an equivalent feedforward network of depth equal to two that effects *exactly* the same computation. The implication of this result is that we need more elaborate information about the properties of the neurons before connectomic constraints on the computational ability of such networks can be discerned.

Before we can examine these questions, we are confronted with the problem of having to define what computation exactly means, in this context. Physically, neurons and their networks are simply devices that receive spike-trains as input, and in turn generate spike-trains as output. It is this translation from spike-trains to spike-trains that characterizes information processing and indeed even cognition in the brain. It is tempting to view a feedforward network as a *transformation*, which is to say a function, that associates a *unique* output spike train with each combination of afferent input spike trains, since such networks do not have recurrent loops. This is the intuition we will seek to make precise.

Since the functional role of single neurons and small networks in the brain is not yet well understood, we do not make assumptions about particular high-level tasks that the network is trying to perform; we are just interested in physical spike-train to spike-train transformations. Likewise, since the kinds of neural code employed are unclear, we make no overarching assumptions about the neural code either. We study precise spike times since there is widespread evidence (Strehler and Lestienne, 1986; Rieke et al, 1997, & references therein) that precise spike times play a role in information processing in the brain, in many cases. Indeed, Spike-Timing Dependent Plasticity, a class of Hebbian learning rules that are sensitive to the relative timing of pre and postsynaptic spikes have been discovered (Markram et al, 1997; Bi and Poo, 1998) that support the role of precise spike-timing in computation in the brain. Studying spike times also subsumes cases where spiking rate may be the relevant parameter and therefore there is no loss of generality in making this assumption.

2 Notation and Preliminaries

In this section, we define the mathematical formalism used to describe spike-trains and frequently-used operations on them that, for instance, shift and segment them. The reader may skim these on the first reading and revisit them if a specific technical point needs clarification later on.

An *action potential* or *spike* is a stereotypical event characterized by the time instant at which it is initiated in the neuron, which is referred to as its *spike time*. Spike times are represented relative to the present by real numbers, with positive values denoting past spike times and negative values denoting future spike times. A *spike-train* $\mathbf{x} = \langle x^1, x^2, \ldots, x^k, \ldots \rangle$ is a strictly increasing sequence of spike times, with every pair of spike times being at least α apart, where $\alpha > 0$ is the absolute refractory period[1] and

[1] We assume a single fixed absolute refractory period for all neurons, for convenience, although our results would be no different if different neurons had different absolute refractory periods.

x^i is the spike time of spike i. An *empty spike-train*, denoted by ϕ, is one which has no spikes. A *time-bounded spike-train* (with *bound* (a, b)) is one where all spike times lie in the bounded interval (a, b), for some $a, b \in \mathbb{R}$. We use $\mathcal{S}$ to denote the set of all spike trains and $\bar{\mathcal{S}}_{(a,b)}$ to denote the set of all time-bounded spike-trains with bound (a, b). A spike-train is said to have a *gap* in the interval (c, d), if it has no spikes in that time interval. Furthermore, this gap is said to be of *length* $d - c$.

We use the term *spike-train ensemble* to denote a collection of spike-trains. Thus, formally, a *spike-train ensemble* $\chi = \langle \mathbf{x}_1, \ldots, \mathbf{x}_m \rangle$ is a tuple of spike-trains. The *order* of a spike-train ensemble is the number of spike-trains in it. For example, $\chi = \langle \mathbf{x}_1, \ldots, \mathbf{x}_m \rangle$ is a spike-train ensemble of order m. A *time-bounded spike-train ensemble* (with *bound* (a, b)) is one in which each of its spike-trains is time-bounded (with *bound* (a, b)). A spike-train ensemble χ is said have a *gap* in the interval (c, d), if each of its spike trains has a gap in the interval (c, d).

Next, we define some operators to time-shift, segment and assemble/disassemble spike-trains from spike-train ensembles. Let $\mathbf{x} = \langle x^1, x^2, \ldots, x^k, \ldots \rangle$ be a spike-train and $\chi = \langle \mathbf{x}_1, \ldots, \mathbf{x}_m \rangle$ be a spike-train ensemble. The *time-shift operator for spike-trains* is used to time-shift all the spikes in a spike-train. Thus, $\sigma_t(\mathbf{x}) = \langle x^1 - t, x^2 - t, \ldots, x^k - t, \ldots \rangle$. The *time-shift operator for spike-train ensembles* is defined as $\sigma_t(\chi) = \langle \sigma_t(\mathbf{x}_1), \ldots, \sigma_t(\mathbf{x}_m) \rangle$. The *truncation operator for spike-trains* is used to "cut out" specific segments of a spike-train. It is defined as follows: $\Xi_{[a,b]}(\mathbf{x})$ is the time-bounded spike-train with bound $[a, b]$ that is identical to $\mathbf{x}$ in the interval $[a, b]$. $\Xi_{(a,b)}(\mathbf{x})$, $\Xi_{(a,b]}(\mathbf{x})$ and $\Xi_{[a,b)}(\mathbf{x})$ are defined likewise. In the same vein, $\Xi_{[a,\infty)}(\mathbf{x})$ is the spike-train that is identical to $\mathbf{x}$ in the interval $[a, \infty)$ and has no spikes in the interval $(-\infty, a)$. Similarly, $\Xi_{(-\infty,b]}(\mathbf{x})$ is the spike-train that is identical to $\mathbf{x}$ in the interval $(-\infty, b]$ and has no spikes in the interval (b, ∞). $\Xi_{(a,\infty)}(\mathbf{x})$ and $\Xi_{(-\infty,b)}(\mathbf{x})$ are also defined similarly. The *truncation operator for spike-train ensembles* is defined as $\Xi_{[a,b]}(\chi) = \langle \Xi_{[a,b]}(\mathbf{x}_1), \ldots, \Xi_{[a,b]}(\mathbf{x}_m) \rangle$. $\Xi_{(a,b)}(\chi)$, $\Xi_{(a,b]}(\chi)$, $\Xi_{[a,b)}(\chi)$, $\Xi_{[a,\infty)}(\chi)$, $\Xi_{(-\infty,b]}(\chi)$, $\Xi_{(a,\infty)}(\chi)$ and $\Xi_{(-\infty,b)}(\chi)$ are defined likewise. Furthermore, $\Xi_t(\cdot)$ is shorthand for $\Xi_{[t,t]}(\cdot)$. The *projection operator for spike-train ensembles* is used to "pull-out" a specific spike-train from a spike-train ensemble. It is defined as $\Pi_i(\chi) = \mathbf{x}_i$, where $1 \leq i \leq m$. Let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ be spike-trains. The *join operator for spike-trains* is used to "bundle-up" a set of spike-trains to obtain a spike-train ensemble. It is defined as $\mathbf{y}_1 \sqcup \mathbf{y}_2 \sqcup \ldots \sqcup \mathbf{y}_n = \bigsqcup_{i=1}^{n} \mathbf{y}_i = \langle \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n \rangle$.

3 The Neuron Model

The present work treats the setting in which we know the wiring diagram of a network, but lack detailed information on the response properties of its neurons. We then wish to show computations that the network cannot accomplish, *no matter what response properties its neurons may have*. The modeling question we must first address, therefore, is what kind of neuron model we ought to use in such a context.

While we lack detailed information on each of the neurons in the network, it is reasonable to assume that all the neurons in the network satisfy a small number of elementary properties. For example, spiking neurons are generally known to have an absolute refractory period and most of them settle to a resting membrane potential upon receiving no input for sufficiently long, where this resting membrane potential is smaller than the threshold required to elicit a spike. We wish to have a model that is contingent on a small number of such basic properties, but whose responses are unconstrained otherwise, in order to allow for a large class of possible responses.

Mathematically, we formulate the neuron as an abstract mathematical object that satisfies a small number of axioms, which correspond to such elementary properties.

Another way to think about the model is as one that brings "under its umbrella" several other neuron models. These are models that satisfy the properties that our model is contingent on. In Online Resource A, we demonstrate, for instance, that neuron models such as the Leaky Integrate-and-Fire Model and the Spike Response Model SRM$_0$ satisfy these properties up to arbitrary accuracy. Our model can thus be seen as a generalization[2] of these neuron models, specifically one that allows for a much wider class of responses.

There are also other strong reasons for employing this type of model. Crucially, it allows the possibility of incrementally adding more properties to the neuron model, and studying how that further constrains the computational properties of the network. This would model the scenario where we have more detailed knowledge about individual neuron properties, which might well turn out to be the case with the connectome projects. While technical hurdles presently lie in the way of inferring, for example, distributions of ion-channels and neurotransmitter receptors in each neuron using electron microscopy(Denk et al, 2012), it is conceivable that future advances make this possible, giving us a better sense of the physiological properties of all the individual neurons in the connectome; other future technological ad-

[2] Models such as the Leaky Integrate-and-Fire (LIF) and Spike Response Model (SRM), in addition to the constraints in our model have their membrane potential function $P(\cdot)$ specified outright. In case of the LIF model, this is specified via a differential equation and in the case of SRM, the specific functional form is written down explicitly.

vances may also help in this direction. Furthermore, the need for adding more properties to the model and studying the consequences will become especially apparent towards the end of this paper, when we show limits to the constraints imposed by the present set of properties assumed in the model.

3.1 Properties

We start off by informally describing the properties that our model is contingent on. Notable cases where the properties do not hold are also pointed out. This is followed by a formal mathematical definition of the model. The approach taken here in defining the model is along the lines of the one in (Banerjee, 2001).

The following are our assumptions:

1. We assume that the neuron is a device that receives input from other neurons exclusively by spikes which are received via chemical synapses.[3]
2. The neuron is a finite-precision device with fading memory. Hence, the underlying potential function can be determined[4] from a bounded past. That is, we assume that, for each neuron, there exist positive real numbers Υ and ρ, so that the current membrane potential of the neuron can be determined as a function of the input spikes received in the past Υ milliseconds and the spikes produced by the neuron in the past ρ milliseconds. The parameter Υ would correspond to the timescale at which the neuron integrates inputs received from other neurons and ρ corresponds to the notion of *relative refractory period.*
3. Specifically, we assume that the membrane potential of the neuron can be written down as a real-valued, everywhere-bounded function of the form $P(\chi; \mathbf{x}_0)$, where $\mathbf{x}_0$ is a time-bounded spike-train, with bound $(0, \rho)$ and $\chi = \langle \mathbf{x}_1, \ldots, \mathbf{x}_m \rangle$ is a time-bounded spike-train ensemble with bound $(0, \Upsilon)$. Informally, $\mathbf{x}_i$, for $1 \leq i \leq m$, is the sequence of spikes afferent in synapse i in the past Υ milliseconds and $\mathbf{x}_0$ is the sequence of spikes efferent from the current neuron in the past ρ milliseconds. The function $P(\cdot)$ characterizes the entire spatiotemporal response of the neuron to spikes including synaptic strengths, their location on dendrites, and their modulation of each other's effects at the soma, spike-propagation delays, and the postspike hyperpolarization.
4. Without loss of generality, we assume the resting membrane potential to be 0.
5. Let $\tau > 0$ be the threshold that the membrane potential must reach in order to elicit a spike. Observe that the model allows for variable[5] thresholds, as long as the threshold itself is a function of spikes afferent in the past Υ milliseconds and spikes efferent from the present neuron in the past ρ milliseconds. Furthermore, when a new output spike is produced, in the model, the membrane potential immediately goes below threshold. That is, the membrane potential function in the model takes values that are at most that of the threshold. This simplifies our condition for an output spike to be that the $P(\cdot)$ merely hits threshold, without having to check if it hits it from below, since it cannot hit it from above. Again, this is done without loss of generality. Additionally, let λ be a negative real number that represents a lower-bound on the values that the membrane potential can take.
6. Output spikes in the recent past tend to have an inhibitory effect, in the following sense[6]:
   $P(\chi; \mathbf{x}_0) \leq P(\chi; \boldsymbol{\phi})$, for all "legal" χ and $\mathbf{x}_0$.
   Thus, our model allows for a wide variety of AHPs. Indeed, the only constraint on AHPs is the one given above. That is, suppose, in the first case that at a certain point in time the neuron received spikes in the past Υ seconds present in χ as input and did not output any spikes in the past ρ milliseconds. In the second case, suppose that at a certain point in time the neuron again received spikes in the past Υ seconds present in χ as input but output some spikes in the past ρ milliseconds. The condition states that the membrane potential in the second case must be at most that of the value in the first case. Thus, our results will be true for any neuron model that has an AHP that obeys this condition.
7. Owing to the absolute refractory period $\alpha > 0$, no two input or output spikes can occur closer than α. That is, suppose $\mathbf{x}_0 = \langle x_0^1, x_0^2, \ldots, x_0^k \rangle$, where $x_0^1 < \alpha$. Then $P(\chi; \mathbf{x}_0) < \tau$, for all "legal" χ.
8. Finally, on receiving no input spikes in the past Υ milliseconds and no output spikes in the past ρ milliseconds, the neuron settles to its resting potential. That is,
   $P(\langle \boldsymbol{\phi}, \boldsymbol{\phi}, \ldots, \boldsymbol{\phi} \rangle; \boldsymbol{\phi}) = 0$.

A *feedforward network of neurons*, is a Directed Acyclic Graph where each vertex corresponds to an instantiation of the neuron model, with the exception of some vertices, designated as input vertices (which are placeholders for

[3] In this work, we do not treat electrical synapses or ephaptic interactions (Shepherd, 2004).

[4] We do not treat stochastic variability in the responses of neurons or neuromodulation in this paper.

[5] In many biological neurons, the membrane potential that the soma (or axon initial segment) must reach, in order to elicit a spike is not fixed at all times and is, for example, a function of the inactivation levels of the voltage-gated Sodium channels. Our model can accomodate this phenomenon, to the extent that this threshold itself is a function of spikes afferent in the past Υ milliseconds and spikes efferent from the present neuron in the past ρ milliseconds.

[6] This is violated, notably, in neurons that have a post-inhibitory rebound.

input spike-trains); one neuron is designated the output neuron. The *order* of a feedforward network is equal to the number of its input vertices. The *depth* of a feedforward network is the length of the longest path from an input vertex to the output vertex.

Next, we formalize the above notions into a rigorous definition of a neuron as an abstract mathematical object.

Definition 1 (Neuron)**.** A *neuron* N is a 7-tuple $\langle \alpha, \Upsilon, \rho, \tau, \lambda, m, P : \bar{\mathcal{S}}^m_{(0,\Upsilon)} \times \bar{\mathcal{S}}_{(0,\rho)} \rightarrow [\lambda, \tau] \rangle$, where $\alpha, \Upsilon, \rho, \tau \in \mathbb{R}^+$ with $\rho \geq \alpha$, $\lambda \in \mathbb{R}^-$ and $m \in \mathbb{Z}^+$. Furthermore,

1. If $\mathbf{x}_0 = \langle x_0^1, x_0^2, \ldots, x_0^k \rangle$ with $x_0^1 < \alpha$, then $P(\chi; \mathbf{x}_0) < \tau$, for all $\chi \in \bar{\mathcal{S}}^m_{(0,\Upsilon)}$ and for all $\mathbf{x}_0 \in \bar{\mathcal{S}}_{(0,\rho)}$.
2. $P(\chi; \mathbf{x}_0) \leq P(\chi; \boldsymbol{\phi})$, for all $\chi \in \bar{\mathcal{S}}^m_{(0,\Upsilon)}$ and for all $\mathbf{x}_0 \in \bar{\mathcal{S}}_{(0,\rho)}$.
3. $P(\langle \boldsymbol{\phi}, \boldsymbol{\phi}, \ldots, \boldsymbol{\phi} \rangle; \boldsymbol{\phi}) = 0$.

A neuron is said to *generate a spike* whenever $P(\cdot) = \tau$.

4 Feedforward Networks as Input-to-Output transformations

As discussed earlier, it is intuitively appealing to view feedforward networks of neurons as transformations that map input spike-trains to output spike-trains. In this section, we seek to make this notion precise by clarifying in what sense, if at all, these networks constitute the said transformations. It will turn out that even single neurons cannot correctly be viewed as such transformations, in general. In the next section, however, we show that under biologically-relevant spiking regimes, we can salvage this view of feedforward networks as spike-train to spike-train transformations.

Let us first consider the simplest type of feedforward network, namely a single neuron. Observe that our abstract neuron model does not explicitly prescribe an output spike-train for a given input spike-train ensemble. That is, recall from the previous section, that the membrane potential of the neuron depends not only on the input spikes received in the past Υ milliseconds, it also depends on the output spikes produced by it in the past ρ milliseconds. Therefore, knowledge of just input spike times in the past Υ milliseconds does not uniquely determine the current membrane potential (and therefore the output spike-train produced from it). It might be tempting to then somehow use the fact that past output spikes are themselves a function of input and output received in the more distant past, and attempt to make the current membrane potential a function of a bounded albeit larger "window" of past input spikes alone. The simple counterexample described in Figure 1 shows that this does not work. In particular, if we attempt to characterize the current membrane potential of the neuron as a function of past input spikes alone, the current membrane potential may depend on the position of an input spike that has occurred arbitrarily long time ago in the past. To sum up, this counterexample proves that, without further restrictions, even a single neuron cannot be correctly viewed as a bounded-length spike-train to spike-train transformation.

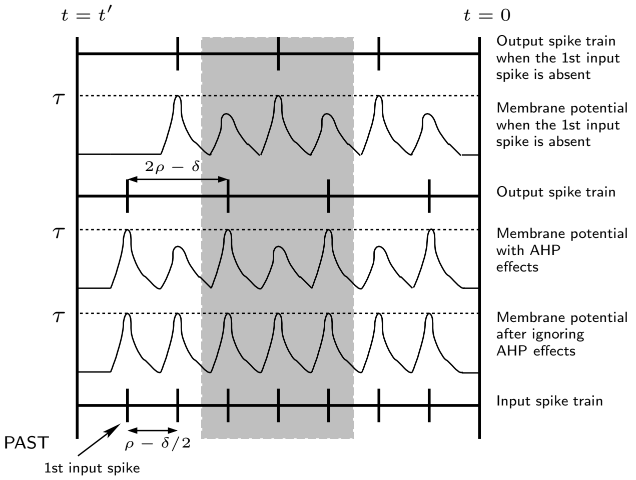

$t = t'$
$t = 0$
Output spike train when the 1st input spike is absent
τ
Membrane potential when the 1st input spike is absent
$2\rho - \delta$
Output spike train
τ
Membrane potential with AHP effects
τ
Membrane potential after ignoring AHP effects
Input spike train
PAST
$\rho - \delta/2$
1st input spike

Fig. 1 This counterexample describes a single neuron which has just one afferent synapse. Until time t' in the past, it received no input spikes. After this time, its input consisted of spikes that arrived every $\rho - \delta/2$ milliseconds, where $0 < \delta \leq 2(\rho - \alpha)$. An input spike alone (if there were no output spikes in the past ρ milliseconds) causes this neuron to produce an output spike. However, in addition, if there were an output spike within the past ρ milliseconds, the afterhyperpolarization (AHP) due to that spike is sufficient to bring the potential below threshold, so that the neuron does not spike currently. We therefore observe that if the first spike of the input spike-train is absent, then the output spike-train changes drastically. Note that this change occurs no matter how often the shaded segment in the middle is replicated, i.e. it does not depend on how long ago the first spike occurred. Thus, the counterexample demonstrates that the membrane potential at any point in time may depend on the position of an input spike that occurred arbitrarily long time ago. Note that the input or the output pattern being periodic and the two output patterns being phase-shifted is not a necessary ingredient of the counterexample; i.e. it is straightforward to construct a (more complicated) counterexample that exhibits this same phenomenon where neither the input spike-train nor the output spike-train are periodic and where the two output spike patterns are not phase-shifted versions of each other.

This pessimistic prognosis notwithstanding, it may seem that if we knew the infinite history of input spikes received by the neuron, we should be able to uniquely determine its current membrane potential. Unfortunately, the situation turns out to be even more dire – this turns out not to be the case. Before we demonstrate this, we must return to the issue of what it means for a neuron to *produce* an output spike-train when it receives a certain spike-train ensemble as input. That is, suppose the reader had an instantiation of our neuron model, which in this case would mean the values of Υ, ρ and τ and the membrane potential function $P(\cdot)$. Further, suppose the reader were given an input spike-train

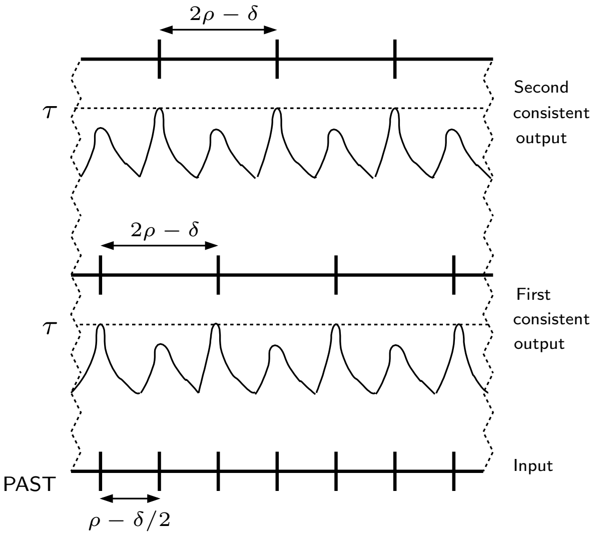


Fig. 2 The counterexample here is very similar to the one in Figure 1, except that, instead of there being no input spikes before t', we have an unbounded input spike-train ensemble, with the same periodic input spikes occurring since the infinite past. The neuron here has the exact same response properties as the one in Figure 1. Observe that both output spike-trains are consistent with this input, for each $t \in \mathbb{R}$. The corresponding membrane potential traces appear below each consistent output spike train.

ensemble χ and told that the neuron "produced" the output spike-train $\mathbf{x}_0$ when driven by χ. Then, all that the reader can do to verify this claim is to check if the given output spike-train is *consistent* with the input spike-train ensemble for the given neuron in the following sense. We would go to each point in time where the neuron spiked and plug into $P(\cdot)$ the input spikes in the past Υ milliseconds from χ, and output spikes from the past ρ milliseconds from $\mathbf{x}_0$ and check if the value of $P(\cdot)$ equals the threshold τ. Likewise, for the time points where the output spike-train does not have a spike, we need to check that this value is less than the threshold. If the answers are in the affirmative for all time-points we can say that the given output spike-train is *consistent* with the given input spike-train ensemble with respect to the neuron in question. However, this still allows the possibility of more than one consistent output spike-train to exist for a given input spike-train ensemble, with respect to a given neuron. Indeed, we will demonstrate that this possibility can occur and therefore given the infinite history of input spikes received by the neuron, we cannot uniquely determine the output spike train produced. Before getting into the counterexample, for completeness, let us formally define this notion of *consistency*. Recall that $\langle t \rangle$ denotes a spike-train with a single spike at time instant t.

Definition 2. An output spike-train $\mathbf{x}_0$ is said to be *consistent* with an input spike-train ensemble χ, with respect to a neuron $\mathsf{N}\langle \alpha, \Upsilon, \rho, \tau, \lambda, m, P : \bar{\mathcal{S}}^m_{(0,\Upsilon)} \times \bar{\mathcal{S}}_{(0,\rho)} \rightarrow [\lambda, \tau] \rangle$, if $\chi \in \mathcal{S}^m$ and the following holds. For every $t \in \mathbb{R}$, $\Xi_t \mathbf{x}_0 = \langle t \rangle$ if and only if $P(\Xi_{(0,\Upsilon)}(\sigma_t(\chi)), \Xi_{(0,\rho)}(\sigma_t(\mathbf{x}_0)) = \tau$.

The question, therefore, is the following. For every (unbounded) input spike-train ensemble χ, does there exist exactly one (unbounded) output spike train $\mathbf{x}_0$, so that $\mathbf{x}_0$ is consistent with χ with respect to a given neuron N? As alluded to, the answer turns out to be in the negative. The counterexample in Figure 2 describes a neuron and an infinitely[7] long input spike-train, which has two consistent output spike-trains.

The underlying difficulty in defining even single neurons as spike-train to spike-train transformations, with both viewpoints discussed above, is persistent dependence, in general, of current membrane potential on "initial state". The way to circumvent this difficulty would be to impose additional restrictions which render such counterexamples untenable. For example, there is the possibility of considering just a subset of input/output spike-trains, which have the property of the current membrane potential being independent of the input spikes beyond a certain time in the past. Such a subset would certainly exclude the examples discussed in this section. This would correspond to restricting our theory to a certain kind of spiking regime.

In the next section, we come up with a condition that, in effect, restricts spike-trains to biologically-relevant spiking regimes and prove that this implies independence as alluded to above. Roughly speaking, the condition is that if a neuron has had a recent gap in its output spike-train equal to at least *twice* its relative refractory period, then its current membrane potential is independent of the input beyond the relatively recent past. We show that this leads to the notion of feedforward networks as spike-train to spike-train transformations to be well-defined.

5 The Gap Lemma and Criteria

In this section, we devise a biologically well-motivated condition that guarantees independence of current membrane potential from input spikes beyond the recent past. This condition is used in constructing a criterion for single neurons which when satisfied, guarantees a unique consistent output spike-train and leads to the view of a neuron as a transformation that maps bounded-length input spike-trains to bounded-length output spike-trains. After this, similar criteria are defined for feedforward networks, in general.

For a neuron, the way input spikes that happened sufficiently earlier affect current membrane potential is via a causal sequence of output spikes, causal in the sense that

[7] The interested reader is referred to Online Resource B for a discussion on the issue of infinitely-long input spike-trains in this context.

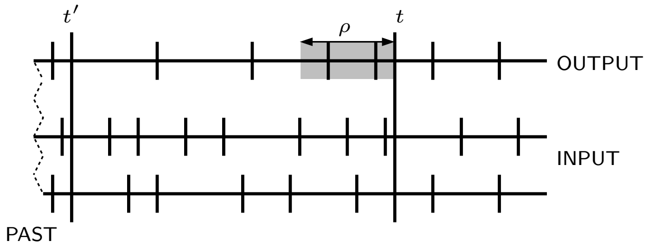


Fig. 3 This figure illustrates the idea behind the Gap Lemma. Suppose there exists a neuron, with Υ and ρ being the lengths of input and output windows respectively, that "effects" the transformation shown above. Let $(t' - t) \geq \Upsilon$. Suppose, the spikes in the shaded region, which is an interval of length ρ occurred at the exact same position, for all input spike-train ensembles that are identical in the range $[t, t']$, but have spikes occurring at arbitrary positions older than time instant t'. Then, the membrane potential of that neuron at t is identical in all those cases. This implies that the spikes in the shaded region are a function of exactly the input spikes in the interval $[t, t']$; in particular, they are independent of input spikes occurring before t'.

each output spike in the sequence had an effect on the membrane potential while the subsequent one in the sequence was being produced and the input spike in question had an effect on the membrane potential, when the oldest output spike in the same sequence was produced. As a result, when an input spike is moved, this effect could propagate across time and cause the output spike train to change drastically. The condition in the Gap Lemma, in effect, seeks to break the causality in this causal chain.

Figure 3 elaborates the main idea behind the condition. Suppose there exists a neuron, with Υ and ρ being the lengths of input and output windows respectively, that "effects" the transformation shown in Figure 3. In a nutshell, if there was a guarantee that spike positions in an interval of length ρ in the output spike train would remain invariant to changes in the past input spike-train ensemble, then this would break the aforementioned causal chain.

The question, of course, is what condition might guarantee such a situation. It turns out that a gap of length 2ρ in the output spike-train suffices, as the next lemma shows. That is, if the neuron effects a transformation with a 2ρ gap, say ending at t, present in the output, then for t' being $\Upsilon + \rho$ milliseconds before t, such that no matter how input spikes older than t' are changed, the latter half of the 2ρ gap is guaranteed to have no spikes in each case. Therefore, membrane potential starting at t, is the same in all such cases. 2ρ also turns out to be the smallest gap length for which this works. Figure 4 offers some brief intuition on why a gap of length 2ρ suffices to guarantee independence. The technical details are in the following lemma. A formal proof is available in Online Resource B.

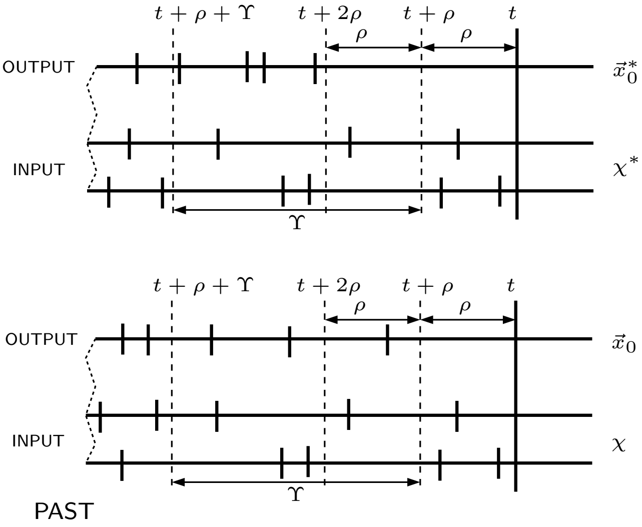


Fig. 4 This figure helps visualize the intuition behind why a gap of length 2ρ suffices to guarantee independence in the Gap Lemma. Suppose a neuron on receiving an input spike-train ensemble χ^* "produces"[8]an output spike-train $\mathbf{x}_0^*$. Further, suppose, $\mathbf{x}_0^*$ has a gap of length 2ρ ending at time instant t. Now let χ be some input spike-train ensemble, which is identical to χ^* in an interval of length $\Upsilon + \rho$ ending at t. Let $\mathbf{x}_0$ be the output spike-train "produced" by χ. Then, the condition guarantees that $\mathbf{x}_0$ has a gap of length ρ immediately preceding t. Here is why. When the neuron is being driven by χ^*, clearly, the membrane potential is below threshold at each time instant ρ milliseconds before t. At each such time instant, the neuron has no past output spikes ρ milliseconds previously. Now, when the neuron is being driven by χ instead, there is no guarantee that the earlier half of the 2ρ gap is preserved . Thus, at each time instant ρ milliseconds before t, the neuron "sees" the same input spike-train ensemble Υ milliseconds previously as with χ^*, but possibly some past output spikes ρ milliseconds previously. Therefore, it's membrane potential at each such time instant may be less than or equal to the corresponding value while the neuron was being driven by χ^*, since, intuitively, the presence of recent efferent spikes could serve to afterhyperpolarize the membrane potential[9]. Thus, since the membrane potential was already below threshold in this time interval while the neuron was being driven by χ^*, it is below the threshold, while the neuron is being driven by χ as well.

Lemma 1 (*Gap Lemma*). *Consider a neuron* $\mathsf{N}\langle \alpha, \Upsilon, \rho, \tau, \lambda, m, P : \bar{\mathcal{S}}^m_{(0,\Upsilon)} \times \bar{\mathcal{S}}_{(0,\rho)} \rightarrow [\lambda, \tau] \rangle$, *a spike-train ensemble* χ^* *of order* m *and a spike-train* $\mathbf{x}_0{}^*$ *which has a gap in the interval* $(t, t + 2\rho)$, *so that* $\mathbf{x}_0{}^*$ *is consistent with* χ^*, *with respect to* N. *Let* χ *be an arbitrary spike-train ensemble that is identical to* χ^* *in the interval* $(t, t + \Upsilon + \rho)$.

Then, every output spike-train consistent with χ, *with respect to* N, *has a gap in the interval* $(t, t + \rho)$. *Furthermore,* 2ρ *is the smallest gap length in* $\mathbf{x}_0^*$, *for which this is true.*

[8] For the sake of simplicity of exposition, assume there is exactly one consistent output spike-train. This is not a requirement as will become clear in the lemma.

[9] Formally, this follows from Axiom 2 in the definition of our abstract neuron.

The Gap Lemma has some ready implications as stated in the corollary below. A proof is available in Online Resource B.

Corollary 1. *Consider a neuron* $\mathsf{N}\langle \alpha, \Upsilon, \rho, \tau, \lambda, m, P : \bar{\mathcal{S}}^m_{(0,\Upsilon)} \times \bar{\mathcal{S}}_{(0,\rho)} \to [\lambda, \tau]\rangle$*, a spike-train ensemble* χ^* *of order* m *and a spike-train* $\mathbf{x}_0{}^*$ *which has a gap in the interval* $(t, t+2\rho)$ *so that* $\mathbf{x}_0{}^*$ *is consistent with* χ^**, with respect to* N*. Then*

1. *Every* $\mathbf{x}_0$ *consistent with* χ^**, with respect to* N*, has a gap in the interval* $(t, t+\rho)$*.*
2. *Every* $\mathbf{x}_0$ *consistent with* χ^**, with respect to* N*, is identical to* $\mathbf{x}_0{}^*$ *in the interval* $(-\infty, t+\rho)$*, i.e. into the future after time instant t* $+ \rho$*.*
3. *For every* t' *more recent than* $(t+\rho)$*, the membrane potential at* t'*, is a function of spikes in* $\Xi_{(t', t+\Upsilon+\rho)}(\chi^*)$*.*

The upshot of the Gap Lemma and its corollary is that whenever a neuron goes through a period of time equal to twice its relative refractory period where it has produced no output spikes it undergoes a "reset" in the sense that its membrane potential from then on becomes independent of input spikes that are older than $\Upsilon + \rho$ milliseconds before the end of the gap.

Large gaps in the output spike-trains of neurons seem to be extensively prevalent in the human brain. In parts of the brain where the neurons spike persistently, such as in the frontal cortex, the spike rate is very low (0.1Hz-10Hz) (Shepherd, 2004). In contrast, the typical spike rate of retinal ganglion cells can be very high but the activity is generally interspersed with large gaps during which no spikes are emitted (Nirenberg et al, 2001).

These observations motivate our definition of a criterion for input spike-train ensembles afferent on single neurons. The criterion stipulates that there be intermittent gaps of length at least twice the relative refractory period in an output spike-train consistent with the input spike-train ensemble, with respect to the neuron in question. As we elaborate in a moment, the definition is set up so that for an input spike-train ensemble χ that satisfies a T-Gap criterion for a neuron, the membrane potential at any point in time is dependent on at most T milliseconds of input spikes in χ before it.

Definition 3 (Gap Criterion for a single neuron)**.** For $T \in \mathbb{R}^+$, a spike-train ensemble χ is said to satisfy a T-Gap Criterion[10] for a neuron $\mathsf{N}\langle \alpha, \Upsilon, \rho, \tau, \lambda, m, P : \bar{\mathcal{S}}^m_{(0,\Upsilon)} \times \bar{\mathcal{S}}_{(0,\rho)} \to [\lambda, \tau]\rangle$ if the following is true: There exists a spike-train $\mathbf{x}_0$ with at least one gap of length 2ρ in every interval of time of length $T - \Upsilon + 2\rho$, so that $\mathbf{x}_0$ is consistent with χ with respect to N.

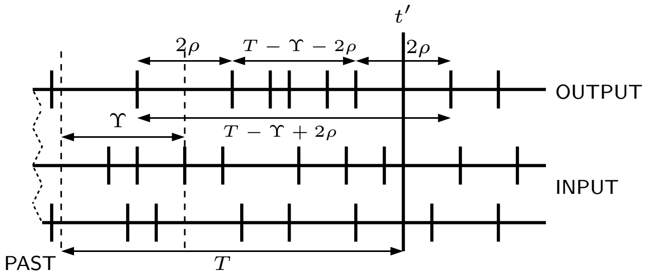


Fig. 5 Illustration demonstrating that for an input spike-train ensemble χ that satisfies a T-Gap criterion, the membrane potential at any point in time is dependent on at most T milliseconds of input spikes in χ before it. Owing to the T-Gap criterion the distance between the end and start of any two consecutive gaps of length 2ρ on the output spike-train is at most $T - \Upsilon - 2\rho$. Upto the earlier half of a 2ρ gap (whose latest point is denoted by t') is dependent on input corresponding to the previous 2ρ gap. It follows that the membrane potential at t' depends only on input spikes in the interval of length T before it, as depicted, owing to the Gap Lemma.

Such input spike-train ensembles also have exactly one consistent output spike-train. The interested reader is directed to Proposition 1 in Online Resource B for a formal statement and proof of this fact.

For an input spike-train ensemble χ that satisfies a T-Gap criterion for a neuron, the membrane potential at any point in time is dependent on at most T milliseconds of input spikes in χ before it, as discussed in Figure 5.

With inputs that satisfy the T-Gap Criterion, here is what we need to do to physically determine the current membrane potential, even if the neuron has been receiving input since the infinite past: Start off the neuron from an arbitrary state, and drive it with input that the neuron received in the past T milliseconds. The Gap Lemma guarantees that the membrane potential we see now will be identical to the actual membrane potential, since the membrane potential is guaranteed to have undergone a "reset" in the ensuing time.

The Gap Criterion we have defined for single neurons can be naturally extended to the case of feedforward networks. The criterion is simply that the input spike-train ensemble to the network is such that every neuron's input obeys a scaled Gap criterion for single neurons. Figure 6 explains the idea. Formally, the definition proceeds inductively, starting with neurons of depth 1.

Definition 4 (Gap Criterion for a feedforward network)**.** An input spike-train ensemble χ is said to satisfy a T-Gap Criterion for a feedforward network if each neuron in the network satisfies a $(\frac{T}{d})$-Gap Criterion, when the network is driven by χ, where d is the depth of the acyclic network.

As with the criterion for the single neuron, the membrane potential of the output neuron at any point is dependent on at most T milliseconds of past input, if the input spike-train ensemble to the feedforward network satisfies

[10] Note that for sufficiently small values of T (in relation to Υ and ρ), no χ may satisfy a T-Gap Criterion. This is deliberate formulation that will minimize notational clutter in forthcoming definitions.

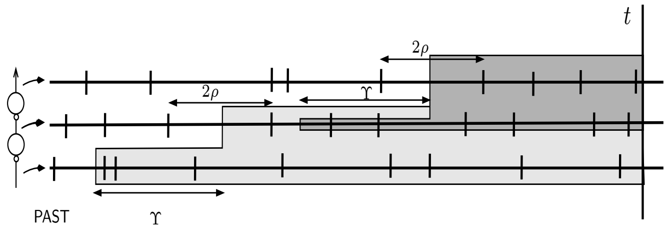


Fig. 6 Schematic diagram illustrating how the Gap criterion works for the simple two-neuron network on the left. The membrane potential of the output neuron at t depends on input received from the "intermediate" neuron, as depicted in the darkly-shaded region, owing to the Gap Lemma. The output of the intermediate neuron in the darkly-shaded region, in turn, depends on input it received in the lightly-shaded region. Thus, transitively, membrane potential of the output neuron at t is dependent at most on input received by the network in the lightly-shaded region.

a T-Gap criterion. Additionally, the output spike-train is unique. Lemma 2 and its proof in Online Resource B make precise these facts.

We thus find ourselves at a juncture where questions we initially sought to ask can be posed in a self-consistent manner. So, looking back at the big picture, we had initially wished to view feedforward networks as transformations that mapped bounded-length input spike-trains to bounded-length output spike trains. However, we found that this notion was not always well-defined. We then showed that if we restrict the set of input spike-trains so they satisfied certain criteria, one can correctly speak of output spike-trains that such inputs are mapped to, by the feedforward network in question. We also argued that this restricted set of spike-trains encompasses biologically-relevant spiking regimes. Thus, feedforward networks can be seen as transformations that map this restricted set of input spike-trains to output spike-trains. Indeed, this will be the sense in which feedforward networks are treated as transformations. Next, we formalize these observations and define some notation.

Notation. Given a feedforward network $\mathcal{N}$, let $\mathcal{G}_{\mathcal{N}}^{T}$ be the set of all input spike-train ensembles that satisfy a T-Gap Criterion for $\mathcal{N}$. Let $\mathcal{G}_{\mathcal{N}} = \bigcup_{T \in \mathbb{R}^+} \mathcal{G}_{\mathcal{N}}^{T}$. Therefore, every feedforward network $\mathcal{N}$ induces a transformation $\mathcal{T}_{\mathcal{N}} : \mathcal{G}_{\mathcal{N}} \rightarrow \mathcal{S}$ that maps each spike-train ensemble in $\mathcal{G}_{\mathcal{N}}$ to a unique output spike train in the set of spike-trains $\mathcal{S}$. Suppose $\mathcal{G}' \subseteq \mathcal{G}_{\mathcal{N}}$. Then, let $\mathcal{T}_{\mathcal{N}}|_{\mathcal{G}'} : \mathcal{G}' \rightarrow \mathcal{S}$ be the map defined as $\mathcal{T}_{\mathcal{N}}|_{\mathcal{G}'}(\chi) = \mathcal{T}_{\mathcal{N}}(\chi)$, for all $\chi \in \mathcal{G}'$.

The Gap Criteria are very general and biologically well-motivated. However, given a neuron or a feedforward network, there does not appear to be an easy way to characterize all the input spike-train ensembles that satisfy a certain Gap Criterion for it. That is, for a given neuron, whether an input spike-train ensemble satisfies a Gap Criterion for it seems to depend intimately on the exact form of its membrane potential function. As a result, a spike-train ensemble that satisfies a Gap criterion for one neuron may not satisfy any Gap Criterion for another neuron. For a feedforward network, the problem becomes even more difficult, since intermediate neurons must satisfy Gap Criteria, and also produce output spike-trains that satisfy Gap Criteria for neurons further downstream. Furthermore, in order to compare transformations effected by two different networks, we need to study inputs that satisfy some Gap criterion for both of them, for otherwise, the notion of a transformation may no longer hold. Now, we sought to ask what transformations *all* feedforward networks with a certain architecture could not do. For this, we need to characterize inputs that satisfy a Gap Criterion for all the networks involved, which seems to be an even more intractable problem.

This brings up the question of the existence of another criterion according to which the set of spike-train ensembles is easier to characterize and is *common* across different networks. Next, we propose one such criterion and show that it consists of spike-train ensembles which are a subset of those induced by the Gap criteria for all feedforward networks. Loosely speaking, these are input spike-train ensembles which, before a certain time instant in the past, have had no spikes. The spike-train ensembles satisfying the said criterion, which we call the Flush criterion, allow us to sidestep the difficult issues just discussed. While this is a purely theoretical construct with no claim of biological relevance, in Section 7, we prove that there is no loss by restricting ourselves to the Flush criterion. That is, not only is a result proved using the Flush criterion applicable with the Gap criterion, *every* result true with the Gap criterion can be proved by using the Flush criterion exclusively.

6 Flush Criterion

The idea of the Flush Criterion is to force the neuron to produce no output spikes for sufficiently long so as to guarantee that a Gap criterion is being satisfied. This is done by having a semi-infinitely long interval with no input spikes. This "flushes" the neuron by bringing it to the resting potential and keeps it there for a sufficiently long time, during which it produces no output spikes. In a feedforward network, the flush is propagated so that all neurons have had a sufficiently long gap in their output spike-trains. Observe that the Flush Criterion is not defined with reference to any feedforward network and is just a property of the spike-train ensemble. We make this notion precise below.

Definition 5 (Flush Criterion)**.** A spike-train ensemble χ is said to satisfy a *T-Flush Criterion*, if all its spikes lie in the interval $(0, T)$, i.e. it has no spikes upto time instant T and since time instant 0.

It turns out that an input spike-train ensemble to a neuron that satisfies a Flush criterion also satisfies a Gap criterion. The technical details along with a proof are in Lemma 3 in Online Resource B.

Likewise, an input spike-train ensemble to a feedforward network satisfying a Flush criterion also satisfies a Gap criterion for that network, as elaborated in Lemma 4 which is available in Online Resource B with a proof.

The Flush criterion is a construct made for mathematical expedience and prima facie does not have any biological relevance. It is a network-independent criterion which enables us to circumvent difficulties that working with the Gap criterion entailed. It will soon become clear why it is a useful construction, when we show that it is equivalent to the Gap criterion insofar as the questions we seek to ask are concerned.

7 Transformational Complexity

Having laid the groundwork, in this section, we set up a definition that will allow us to ask if there exists a transformation that no network of a certain architecture could effect that a network of a different architecture could. It is convenient to formulate the definition in the following terms. Given two classes[11] of networks with the second class encompassing the first, we ask if there is a network in the second class whose transformation cannot be performed by any network in the first class. That is, does the second class possess a larger repertoire of transformations than the first, giving it *more complex* computational capabilities?

Definition 6 (Transformational Complexity)**.** Let Σ_1 and Σ_2 be two sets of feedforward networks, each network being of order m, with $\Sigma_1 \subseteq \Sigma_2$. Define $\mathcal{G}_{12} = \bigcap_{\mathcal{N} \in \Sigma_2} \mathcal{G}_{\mathcal{N}}$. The set Σ_2 is said to be *more complex than* Σ_1, if there exists an $\mathcal{N}' \in \Sigma_2$ such that for all $\mathcal{N} \in \Sigma_1, \mathcal{T}_{\mathcal{N}'}|_{\mathcal{G}_{12}} \neq \mathcal{T}_{\mathcal{N}}|_{\mathcal{G}_{12}}$.

A couple of remarks about the definition above are in order. Firstly, Σ_1 being a proper subset of Σ_2, does not necessarily imply that the that the set of transformations effected by networks in Σ_1 is also a proper subset of those effected by Σ_2. In particular, it could be the case that the set of transformations effected by Σ_1 is exactly the same as that effected by Σ_2, even though Σ_1 is a proper subset of Σ_2. Indeed, this is what is demonstrated by the result of Section 9, which shows in the context of the present neuron model that even though the set of depth-two feedforward networks is a strict subset of the set of all feedforward networks, both these sets effect the same class of transformations, namely those that are causal, time-invariant and resettable. Secondly, observe that while comparing a set of networks, we restrict ourselves to inputs for which all the networks satisfy a certain Gap Criterion (though, not necessarily for the same T), so that the notion of a transformation is well-defined on the input set, for all networks under consideration. Note also that $\mathcal{G}_{12}$ is always a nonempty set, because $\mathcal{G}_{12}$ contains within it all inputs satisfying the Flush criterion. Henceforth, for brevity, any result that establishes a relationship of the form defined above is called a *complexity result*. Before we proceed, we introduce some useful notation.

Notation. Let the set of spike-train ensembles of order m that satisfy the T-Flush criterion be $\mathcal{F}_m^T$. Let $\mathcal{F}_m = \bigcup_{T \in \mathbb{R}^+} \mathcal{F}_m^T$. What we have established in the previous section is that $\mathcal{F}_m \subseteq \mathcal{G}_{\mathcal{N}}$, for every feedforward network $\mathcal{N}$ of order m.

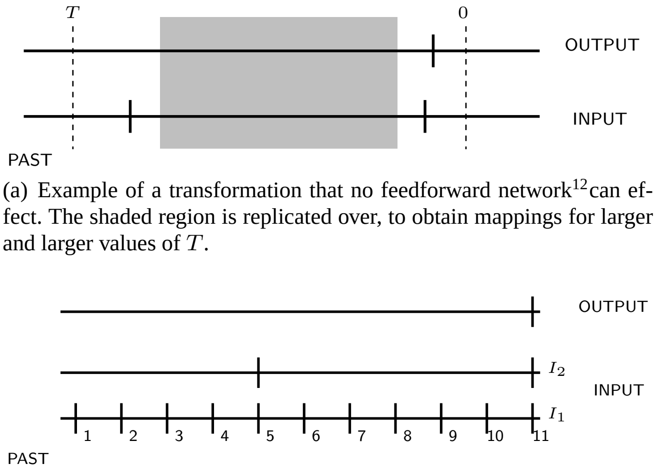
Next, we show that if one class of networks is more complex than another, then inputs that satisfy the Flush Criterion are both necessary and sufficient to prove this. That is, to prove this type of complexity result, one can work exclusively with Flush inputs without losing any generality. This is not obvious because Flush inputs form a subset of the more biologically well-motivated Gap inputs. The next lemma formalizes this equivalence. Note that the statement of the lemma is substantially identical to that of Definition 6, except that the input spike-train ensembles in the lemma below satisfy the Flush criterion, as opposed to the ones in Definition 6 which satisfy $\mathcal{G}_{12}$, the set of input spike-train ensembles that satisfy a Gap Criterion for all the networks under consideration.

Lemma 5 (Equivalence of Flush and Gap Criteria with respect to Transformational Complexity)**.** *Let Σ_1 and Σ_2 be two sets of feedforward networks, each network being of order m, with $\Sigma_1 \subseteq \Sigma_2$. Then, Σ_2 is more complex than Σ_1 if and only if $\exists \mathcal{N}' \in \Sigma_2$ such that $\forall \mathcal{N} \in \Sigma_1, \mathcal{T}_{\mathcal{N}'}|_{\mathcal{F}_m} \neq \mathcal{T}_{\mathcal{N}}|_{\mathcal{F}_m}$.*

Proof sketch. A full proof is available in Online Resource B; here we sketch the intuition behind the proof.

Showing that Flush inputs are sufficient is the easier half of the proof. If a complexity result can be shown using Flush inputs, it follows that it holds for Gap inputs as well, since $\mathcal{F}_m \subseteq \mathcal{G}_{12}$. To show that the existence of Flush inputs is necessary, we assume a complexity result proved using Gap inputs and construct Flush inputs such that the result can be shown using those Flush inputs alone. Now suppose $\mathcal{N}' \in \Sigma_2$ be the network such that no network in Σ_1 effects the same transformation as $\mathcal{N}'$, when the domain is restricted to the set $\mathcal{G}_{12}$. Now, consider arbitrary $\mathcal{N} \in \Sigma_1$. There must exist a $\chi \in \mathcal{G}_{12}$ such that $\mathcal{T}_{\mathcal{N}'}|_{\mathcal{F}_m}(\chi) \neq \mathcal{T}_{\mathcal{N}}|_{\mathcal{F}_m}(\chi)$. By definition, this χ satisfies a T_1-Gap Criterion for $\mathcal{N}$ and a T_2-Gap Criterion for $\mathcal{N}'$. Let $T = \max(T_1, T_2)$. The claim is that if χ is cut up into "chunks" of length $2T$, where each "chunk" satisfies a 2T-Flush criterion, then $\mathcal{N}$ and $\mathcal{N}'$ will map at least one of those chunks to different output spike

[11] The classes of networks could correspond to ones that contain all networks with specific network architectures, although for the purpose of the definition, there is no reason to require this to be the case.



(a) Example of a transformation that no feedforward network[12]can effect. The shaded region is replicated over, to obtain mappings for larger and larger values of T.

(b) A transformation that no single neuron can effect, that a network with two neurons can.

Fig. 7

trains, since the output in the latter half of the chunk is identical to that produced by the corresponding segment of χ. This process of "cutting up", when "completed" for each $\mathcal{N} \in \Sigma_1$ yields a subset of Flush inputs, using which the complexity result can be established. □

Assured by this theoretical guarantee that there is no loss of generality by doing so, we will henceforth only work with inputs satisfying the Flush Criterion, while faced with the task of proving complexity results. This buys us a great deal of mathematical expedience at no cost. From now on, unless qualified otherwise, when we speak of a *transformation*, we mean a map of the form $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$ that maps the set of Flush input spike-train ensembles to the set of output spike-trains.

8 Complexity results

In this section, we establish some complexity results. First, we show that there exist spike-train to spike-train transformations that no feedforward network can effect. Next, we show a transformation that no single neuron can effect but a network consisting of two neurons can. After this, we prove a result which shows that a class of architectures that share a certain structural 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. The interested reader is directed to Online Resource B for some technical remarks concerning the mechanics of proving complexity results that are not central to the exposition here.

Before establishing complexity results, we point out that it is straightforward to construct a transformation that cannot be effected by any feedforward network. One of its input spike-train ensembles with the prescribed output is shown in Figure 7(a). For larger T, the shaded region is simply replicated over and over again. Informally, the reason this transformation cannot be effected by any network is that, for any network, beyond a certain value of T, the shaded region tends to act as a "flush", erasing "memory" of the first input spike. When the network receives another input spike, it is in the exact same "state" it was when it received the first input spike, and therefore cannot produce an output spike after the second input spike.

Next, we prove that the set of feedforward networks with at most two neurons is more complex than the set of single neurons. The proof is by prescribing a transformation which cannot be done by any single neuron. We then construct a network with two neurons that can indeed effect this transformation. Note that in the statement of the theorem below, m stands for the number of input spike trains.

Theorem 1. *Suppose $m \geq 2$. Let Σ be the set of feedforward networks with at most two neurons that each receive an input spike-train ensemble of order m. Then, Σ is more complex than the set of single neurons of order m.*

Proof. We first prescribe a transformation, prove that it cannot be effected by a single neuron and then construct a two-neuron network and show that it can indeed effect the same transformation.

We first prove the result for $m = 2$ and later indicate how it can be extended for larger values of m. Let the two input spike-trains in each input spike-train ensemble, which satisfies a Flush Criterion be I_1 and I_2. Figure 7(b) illustrates the transformation. Informally, I_1 has regularly-spaced spikes starting after time instant T until 0. I_2 has two spikes, with the first one, loosely speaking, in the "middle" of $(0, T)$ and the second one at the end, i.e. right before time instant 0. An output spike is always prescribed after the second spike in I_2 occurs, and not elsewhere. For larger T, the number of spikes on I_1 increases so as to maintain the same regular spacing; I_2, in contrast, still has just two spikes, the first one roughly in the middle and the second in the end. For the sake of exposition, we call the distance between consecutive spikes on I_1, one time unit and we number the spikes of I_1 with the first spike being the oldest one.

More precisely, the transformation is prescribed for a subset of $\mathcal{F}_m$, whose elements are indexed by $i = 1, 2, \cdots$. Figure 7(b) illustrates the transformation, for $i = 2$. The ith input spike-train ensemble in this subset satisfies a T-Flush criterion, where $T = 4i + 3$ time units. In the ith spike-train ensemble, I_2 has spikes at time instants at which spike numbers $2i + 1$ and $4i + 3$ occur in I_1. Finally, the output spike-train corresponding to the ith input spike-train ensem-

[12] Recall that the neurons considered in this work are deterministic.

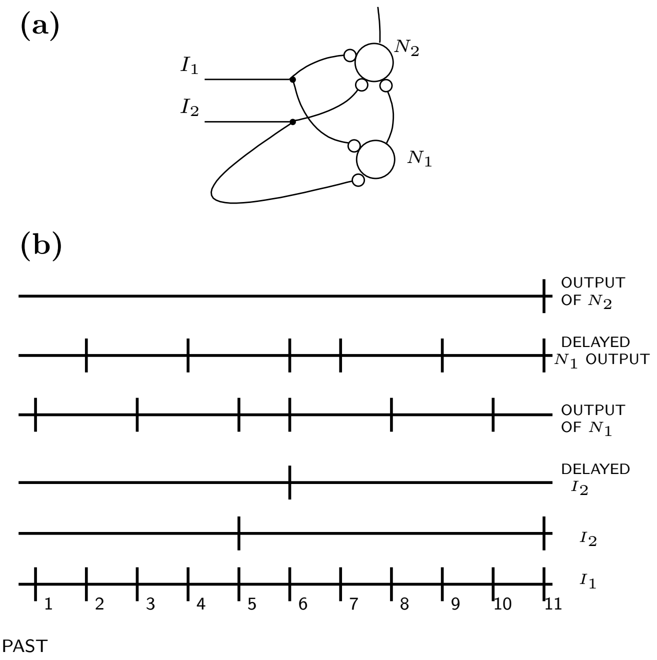


Fig. 8 (a) The network that can effect the transformation described in Figure 7(b). (b) Figure describing the operation of this network.

ble has exactly one spike after[13] the time instant at which I_1 has spike number $4i + 3$.

Next, we prove that the transformation prescribed above cannot be effected by any single neuron. For the sake of contradiction, suppose it can, by a neuron with associated Υ and ρ. Let $\max(\Upsilon, \rho)$ be bounded from above by k time units. We show that for $i \geq \lceil \frac{k}{2} \rceil$, the ith input spike-train ensemble cannot be mapped by this neuron to the prescribed output spike train. For $i = \lceil \frac{k}{2} \rceil$, consider the membrane potential of the neuron after the time instants corresponding to the $(k + 1)$th spike number and $(2k + 3)$rd spike number of I_1. At each of these corresponding time instants, the input received in the past k time units and the output produced by the neuron in the past k time units are the same. Therefore, the neuron's membrane potential must be identical as well. However, the transformation prescribes no spike in one of the first time instants and a spike in the second, which is a contradiction. It follows that no single neuron can effect the prescribed transformation.

We now construct a two-neuron network which can carry out the prescribed transformation. The network is shown in Figure 8(a). I_1 and I_2 arrive instantaneously at N_2. I_1 arrives instantaneously at N_1 but I_2 arrives at N_1 after a delay of 1 time unit. Spikes output by N_1 take one time unit to arrive at N_2, which is the output neuron of the network. The functioning of this network for $i = 2$ is described in Figure 8(b). The generalization for larger i is straightforward. All inputs are excitatory. N_1 is akin to the neuron described in Figure 1, in that while the depolarization due to a spike in I_1 causes potential to cross threshold, if, additionally, the previous output spike happened one time unit ago, the associated hyperpolarization is sufficient to keep the membrane potential below threshold now. However, if there is a spike from I_2 also at the same time as from I_1, the depolarization is sufficient to cause an output spike, irrespective of if there was an output spike one time unit ago. The Υ corresponding to N_2 is shorter than 1 time unit. Further, N_2 produces a spike if and only if all three of its afferent synapses receive spikes at the same time. In the figure, N_1 spikes after times $1, 3, 5$. It spikes after 6 because it received spikes both from I_1 and I_2 at that time instant. Subsequently, it spikes after 8 and 10. The only time wherein N_2 received spikes at all three synapses at the same time is at 11, after which is the prescribed time for the output spike. The generalization for larger i is straightforward.

For larger m, to construct a transformation that cannot be done by a single neuron but can be, by a two-neuron network, one can just have the same input as I_1 or I_2 on the extra input spike trains and the same proof generalizes easily. □

The previous result might seem to suggest that the more the number of neurons (and connections between them) the larger the variety of transformations possible. The next complexity result demonstrates, on the contrary, that the structure of the network architecture is crucial. That is, we can construct network architectures with arbitrarily large number of neurons which cannot perform transformations that a two-neuron network with simple neurons can.

First, we define the structural property that characterizes this class of architectures.

Definition 7 (Path-plural Network)**.** A feedforward network of order m is called *path-plural* if for every set of m paths, where the ith path starts at ith input vertex and ends at the output vertex, the intersection of the m paths is exactly the output vertex.

Every feedforward network in which all the inputs aren't afferent on every neuron, must have embedded within it a path-plural network. For this reason, path-plural networks are an important and ubiquitous class of feedforward networks. How large such networks are in the brain remains to be seen, and this will become clearer as we get more and more data from the connectomics efforts. But, it is conceivable that such networks exist in feedforward pathways that that converge onto networks that, for example, integrate information from multiple sensory modalities.
We now state and prove the complexity result.

Theorem 2. *For $m \geq 3$, let Σ_1 be the set of all path-plural feedforward networks of order m. Let Σ_2 be the union of Σ_1*

[13] Strictly speaking, the output spike happens at $4i + 3 + \epsilon$, where $\epsilon > 0$ is a small real number. Henceforth whenever we say an output spike is *after* a certain time instant, we mean it in this sense.

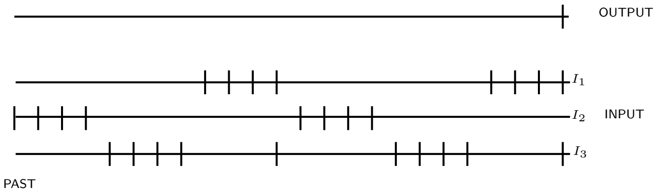


Fig. 9 A transformation that no feedforward network of order 3 with a path-plural architecture can effect.

with the set of all two-neuron feedforward networks of order m. Then, Σ_2 is more complex than Σ_1.

Proof. We first prescribe a transformation, prove that it cannot be effected by any network in Σ_1 and then construct a two-neuron network and show that it can indeed effect the same transformation.

We prove the theorem for $m = 3$; the generalization for larger m is straightforward. The following transformation is prescribed for $m = 3$. Let the three input spike-trains in each input spike train ensemble, which satisfies a Flush Criterion be I_1, I_2 and I_3. As before, we will use regularly spaced spikes; we call the distance between two such consecutive spikes one time unit and number these spike time instants with the oldest being numbered 1; we call this numbering the spike index. Again, the transformation is prescribed for a subset of $\mathcal{F}_m$, whose elements are indexed by $i = 1, 2, \cdots$. Figure 9 illustrates the transformation for $i = 2$. The ith input spike-train ensemble in the subset satisfies a T-Flush Criterion for $T = 4im$ time units. The first $2i$ time units have spikes on I_2 spaced one time unit apart, the next $2i$ on I_3 and so forth. In addition, at spike index $2im$, I_m has a single spike. The input spike pattern from the beginning is repeated once again for the latter $2im$ time units. The prescribed output spike-train has exactly one spike after spike index $4im$.

Next we prove that the transformation prescribed above cannot be effected by any network in Σ_1. For the sake of contradiction, assume that there exists a network $\mathcal{N} \in \Sigma_1$ that can effect the transformation. Let Υ and ρ be upper bounds on the same parameters over all of the neurons in $\mathcal{N}$ and let d be the depth of $\mathcal{N}$. By construction of Σ_1, every neuron in $\mathcal{N}$ that is afferent on the output neuron receives input from at most $m - 1$ of the input spike-trains; for, otherwise there would exist a set of m paths, one from each input vertex to the output neuron, whose intersection would contain the neuron in question. The claim, now, is that for $i > \frac{\Upsilon d}{2} + \rho$, the output neuron of $\mathcal{N}$ has the same membrane potential at spike index $2im$ and $4im$, and therefore either has to spike at both those instants or not. Intuitively, this is so because each neuron afferent on the output neuron receives a "flush" at some point after $2im$, so that the output produced by it Υ milliseconds before time index $2im$ and Υ milliseconds before time index $4im$ are the same. This is straightforward to verify.

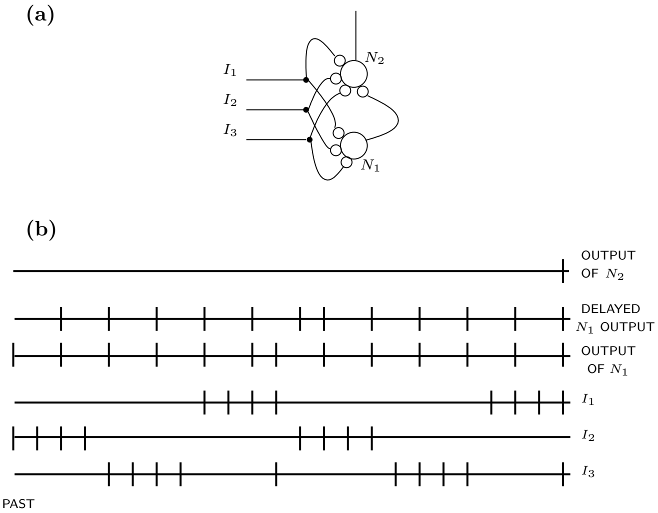


Fig. 10 (a) Network that can effect the transformation described in Figure 9. (b) Figure describing the operation of this network.

We now construct a two-neuron network that can effect this transformation. The construction is similar to the one used in Theorem 1. For $m = 3$, the network is shown in Figure 10. I_1, I_2 and I_3 arrive instantaneously at N_1 and N_2. Spikes output by N_1 take two time units to arrive at N_2, which is the output neuron of the network. The functioning of this network for $i = 2$ is described in Figure 10(b). The generalization for larger i is straightforward. All inputs are excitatory. N_1 is akin to the the neuron N_1 used in the network in Theorem 1 except that that periodic input may arrive from any one of I_1, I_2 or I_3. As before, if two input spikes arrive at the same time, as in spike index $2im$, the depolarization is sufficient to cause an output spike in N_1, irrespective of if there was an output spike one time unit ago. Again, the Υ corresponding to N_2 is shorter than 1 time unit and N_2 produces a spike if and only if three of its afferent synapses receive spikes at the same time instant. As before, the idea is that at time $2im$, N_2, receives two spikes, but not a spike from N_1, since it is "out of sync". However, at time $4im$, additionally, there is a spike from N_1 arriving at N_2, which causes N_2 to spike. □

To conclude, what we have demonstrated in this section is that, for certain classes of networks, just by knowing the architecture of the network, we can rule out computations that the network could be doing. All we assumed was that the neurons in the network satisfy a small number of elementary properties; notably these results do not require knowledge of detailed physiological properties of the neurons in the network. This, in itself, is somewhat surprising due to the intuitively-appealing expectation that network structure may not impose as strong a constraint as neurophysiology insofar as the computational ability of a network is concerned. In the next section, however, we show that this intuition is sound in some cases by proving that there are limits to the constraints imposed by network structure in the presence of very limited information on the physiology.

9 Limits to constraints imposed by network structure

The main thrust of this work, thus far, has been in demonstrating that connectomic constraints do indeed restrict the computational ability of certain networks, even when we do not assume much about the physiological properties of their neurons. As one might expect, we should be able to get better mileage, so to speak, if we had more elaborate information on the response properties of the individual neurons. Conversely, it is logical to expect that there might be fundamental limits to what can be said about the computational properties of networks, given very limited knowledge of the neurophysiology of its neurons. In this section, we prove this to be the case. In particular, we show that the small set of assumptions made about our model neurons lead to the absence of connectomic constraints on computation for the class of feedforward networks of depth equal to two. More precisely, it turns out that there does not exist a transformation that cannot be performed by any network of depth two[14] that in turn can be effected by another network (of a different architecture). What this result implies is that one *needs* to make further assumptions on the properties obeyed by the model neurons, before connectomic constraints on this class of networks appear.

So, how does one prove that there does not exist a transformation that cannot be performed by any network of depth two that in turn can be effected by another network? Equivalently, we need to prove that given an arbitrary feedforward network, there exists a feedforward network of depth two that effects *exactly* the same transformation.

The difficulty in proving that every feedforward network, having arbitrary depth, has an equivalent network of depth two, appears to be in devising a way of "collapsing" the depth of the former network, while keeping the effected transformation the same. Our proof actually does not demonstrate this head-on, but instead proves it to be the case indirectly. The broad attack is the following: Consider the set of transformations spanned by the set of all feedforward networks. Recall that this is a proper subset of the set of all transformations, since we had shown a transformation that no feedforward network could effect. The idea is to start off with a certain "nice" subset of the set of all transformations and show that every transformation effected by feedforward networks certainly lies within this subset. Thereafter, we prove, by providing a construction, that every transformation in this "nice" subset can in fact be effected by a feedforward network of depth two[15]. Together, this implies that, for every transformation that can be effected by a feedforward network, there exists a feedforward network of depth two that can effect exactly that transformation.

The interested reader is directed to Online Resource C, which is a 24-minute video that provides an intuitive outline of the results in this section using animations.

Technical structure of the proof

The main theorem that we prove in this section is the following.

Theorem 3. *If $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$ can be effected by a feedforward network, then it can be effected by a feedforward network of depth two.*

This theorem follows from the following two lemmas which are proved in the two subsections that follow:

Lemma 6. *If $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$ can be effected by a feedforward network, then $\mathcal{T}(\cdot)$ is causal, time-invariant and resettable.*

Lemma 7. *If $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$ is causal, time-invariant and resettable, then it can be effected by a feedforward network of depth two.*

9.1 Causal, Time-Invariant and Resettable Transformations

In this section, we first define notions of causal, time-invariant and resettable transformations[16]. Transformations that are causal, time-invariant and resettable form a strict subset of the set of all transformations. We then show that transformations effected by feedforward networks always lie within this subset. This is the relatively easy part of the proof. The next subsection proves the harder part, namely that every transformation in this subset can indeed be effected by a feedforward network of depth equal to two.

Informally, a *causal transformation* is one whose current output depends only on its past input (and not current or future input). Abstractly, it is convenient to define a causal transformation as one that, given two different inputs that are identical until a certain point in time, also have their outputs, according to the transformation, be identical up to (at least) the same point.

[14] equipped with instances of our model neurons

[15] As a by-product, the proof also ends up providing a complete characterization of the set of transformations spanned by the set of all feedforward networks equipped with neurons of the present abstract model, which turns out to be exactly this "nice" set.

[16] Recall that when we say transformation, without further qualification, we mean one, of the form $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$.

Definition 8 (Causal Transformation)**.** A transformation $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ is said to be *causal* if, for every $\chi_1, \chi_2 \in \mathcal{F}_m$, with $\Xi_{(t,\infty)}\chi_1 = \Xi_{(t,\infty)}\chi_2$, for some $t \in \mathbb{R}$, we have $\Xi_{[t,\infty)}\mathcal{T}(\chi_1) = \Xi_{[t,\infty)}\mathcal{T}(\chi_2)$.

As in signals and systems theory, a *time-invariant transformation* is one which always transforms the time-shifted version of an input, to a time-shifted version of its corresponding output. To keep the definition sound, we also need to ensure that the time-shifted input, in fact, also satisfies the Flush criterion.

Definition 9 (Time-Invariant Transformation)**.** A transformation $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ is said to be *time-invariant* if, for every $\chi \in \mathcal{F}_m$ and every $t \in \mathbb{R}$ with $\sigma_t(\chi) \in \mathcal{F}_m$, we have $\mathcal{T}(\sigma_t(\chi)) = \sigma_t(\mathcal{T}(\chi))$.

A *resettable transformation* is one for which there exists a positive real number W, so that an input gap of the form $(t, t+W)$ "resets" it, i.e. output beyond t is independent of input received before it. Again, abstractly, it becomes convenient to say that the output in this case is identical to that produced by an input which has no spikes before t, but is identical to the present input thereafter.

Definition 10 (W-Resettable Transformation)**.** For $W \in \mathbb{R}^+$, a transformation $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ is said to be W-*resettable* if, for every $\chi \in \mathcal{F}_m$ which has a gap in the interval $(t, t+W)$, for some $t \in \mathbb{R}$, we have $\Xi_{(-\infty,t]}\mathcal{T}(\chi) = \mathcal{T}(\Xi_{(-\infty,t]}\chi)$.

Definition 11 (Resettable Transformation)**.** A transformation $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ is said to be *resettable* if, there exists a $W \in \mathbb{R}^+$, so that it is W-resettable.

Next, we prove that every transformation that can be effected by a feedforward network is causal, time-invariant and resettable, in the context of our neuron model and its assumptions.

Lemma 6. *If $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ can be effected by a feedforward network, then $\mathcal{T}(\cdot)$ is causal, time-invariant and resettable.*

Proof sketch. If $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ can be effected by a single neuron it is relatively straightforward to verify that $\mathcal{T}(\cdot)$ is causal, time-invariant and resettable. That it is causal and time-invariant follows from the fact that the $P(\cdot)$ function of the neuron only "looks" at the recent past and not the present or the future to determine membrane potential. That $\mathcal{T}(\cdot)$ is resettable follows from Axiom (3) of the neuron and the Gap Lemma. For a feedforward network, the proof proceeds by mathematical induction on the depth of the network. A full proof is provided in Online Resource B. □

9.2 Construction of a depth two feedforward network for every causal, time-invariant and resettable transformation

In this subsection, we prove the following lemma.

Lemma 7. *If $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ is causal, time-invariant and resettable, then it can be effected by a feedforward network of depth two.*

Before diving into the proofs, we offer some intuition.

Suppose we had a transformation $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ which is causal, time-invariant and resettable. For the moment, pretend it satisfies the following property: There exist constant-sized input and output "windows" so that, for every input spike-train ensemble satisfying a flush criterion, just given knowledge of spikes in those windows of past input and output, one can unambiguously determine, at any point in time, if the transformation prescribes an output spike or not. Intuitively, it seems reasonable that such a transformation can be effected by a single neuron[17] by setting the Υ and ρ of the neuron to the sizes of the input and output windows mentioned above.

Of course, one easily sees that not every transformation that is causal, time-invariant and resettable satisfies the aforementioned property. That is, there could exist two different input instances, whose past inputs and outputs are identical in the aforementioned windows at some points in time; yet in one instance, the transformation prescribes an output spike, whereas it prescribes none in the other. Indeed, the two input instances must differ at some point in the past, for otherwise the transformation would not be causal. Therefore, in such a situation, it is natural to ask if a single "intermediate" neuron can "break the tie". That is, if two input instances differ at some point in the past, the output of the intermediate neuron since then, in any interval of time of length U, must be different in either case, where U is a fixed constant. This is so that a neuron receiving input from the intermediate neuron can *disambiguate* the two inputs, were an output spike demanded for one input but not the other. Unfortunately, this exact property cannot be achieved by any single "tie-breaker" neuron because every transformation induced by a neuron is resettable. In other words, the problem is that, suppose two input instances differ at a certain point in time; however, since then, both have had an arbitrarily large input gap. The input gap serves to "erase memory" in any network that received it and therefore it cannot disambiguate two inputs beyond this gap. Now, fortunately, it does not have to, since this gap also causes a "reset" in the transformation (which is resettable). That is, if such an arbitrarily large gap were present in the input, the transformation would not afterward demand an output spike in one case and no output spike in another. This is because it is W-resettable and therefore cannot make such demands, for input gaps[18] larger than W. Thus, we can make do with a slightly weaker condition; that the intermediate neuron is only guaranteed to

[17] Strictly speaking, it turns out that this is not true; axiom 2 may be violated.

[18] which we call a "reset gap" from now on, for the sake of exposition.

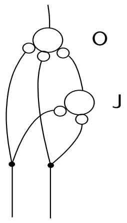


Fig. 11 The network architecture for (order two) feedforward networks of depth two equipped with model neurons described in Section 3 that can effect any causal, time-invariant and resettable transformation.

break the tie, when it is required to do so. That is, suppose there are two input instances, whose outputs according to $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ are different at certain points in time. Then, the corresponding inputs are different too at some point in the past with no reset gaps in the intervening time and therefore the intermediate neuron ought to break the tie. Additionally, for technical reasons that will become clear later, we stipulate that the outputs of the intermediate neuron in the preceding U milliseconds are guaranteed to be different, only if the inputs themselves in the past U milliseconds are not different.

The network we have in mind is illustrated in Figure 11, for $m = 2$. In the following proposition, we prove that if the intermediate neuron satisfies the "tie-breaker" condition alluded to above, then there exists an output neuron, so that the network effects the transformation in question. Thereafter, in the subsequent proposition, we provide a construction for the intermediate neuron that satisfies this condition. By way of notation, recall that $\Xi_0(\cdot)$ is shorthand for $\Xi_{[0,0]}(\cdot)$

Proposition 2. *Let $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ be causal, time-invariant and resettable. Let J be a neuron with $\mathcal{T}_\mathsf{J} : \mathcal{F}_m \to \mathcal{S}$, so that for each $\chi \in \mathcal{F}_m$, $\mathcal{T}_\mathsf{J}(\chi)$ is consistent with χ with respect to J. Further, suppose there exists a $U \in \mathbb{R}^+$ so that for all $t_1, t_2 \in \mathbb{R}$ and $\chi_1, \chi_2 \in \mathcal{F}_m$ with $\Xi_0\sigma_{t_1}(\mathcal{T}(\chi_1)) \neq \Xi_0\sigma_{t_2}(\mathcal{T}(\chi_2))$, we have $\Xi_{(0,U)}(\sigma_{t_1}(\mathcal{T}_\mathsf{J}(\chi_1) \sqcup \chi_1)) \neq \Xi_{(0,U)}(\sigma_{t_2}(\mathcal{T}_\mathsf{J}(\chi_2) \sqcup \chi_2))$.*

Then, there exists a neuron O, so that for every $\chi \in \mathcal{F}_m$, $\mathcal{T}(\chi)$ is consistent with $\mathcal{T}_\mathsf{J}(\chi) \sqcup \chi$ with respect to O.

Proof sketch. The straightforward way for the neuron O to effect $\mathcal{T}(\cdot)$ is to determine the points of time wherein an output spike is prescribed and set its membrane potential function to hit threshold at those instances. Since the neuron J essentially "disambiguates" the input, this assignment can be done without conflict. However, we also need to show that doing this does not violate any of the three axioms of our abstract model, for the neuron O. Axiom (1) follows easily from the fact that the co-domain of $\mathcal{T}(\cdot)$ is $\mathcal{S}$. Axiom (3) takes some work to show and uses the fact that $\mathcal{T}(\cdot)$ is causal, time-invariant and resettable. Axiom (2), on the other hand, presents some subtleties. Now, in addition to setting membrane potential to threshold at the aforementioned points, in order to satisfy Axiom (2), we would also need to set it to hit threshold, when the input window has the same pattern and the output window is empty instead. However, with this assignment, we need to then show that no spurious spikes are generated. This takes a little work and again uses the "tie-breaker" condition of the intermediate neuron J. The full proof is available in Online Resource B. □

The next proposition shows that one can always construct an intermediate neuron that satisfies the said "tie-breaker" condition.

Proposition 3. *Let $\mathcal{T} : \mathcal{F}_m \to \mathcal{S}$ be causal, time-invariant and resettable. Then there exists a neuron J and $U \in \mathbb{R}^+$ so that for all $t_1, t_2 \in \mathbb{R}$ and $\chi_1, \chi_2 \in \mathcal{F}_m$ with $\Xi_0\sigma_{t_1}(\mathcal{T}(\chi_1)) \neq \Xi_0\sigma_{t_2}(\mathcal{T}(\chi_2))$, we have $\Xi_{(0,U)}(\sigma_{t_1}(\mathcal{T}_\mathsf{J}(\chi_1) \sqcup \chi_1)) \neq \Xi_{(0,U)}(\sigma_{t_2}(\mathcal{T}_\mathsf{J}(\chi_2) \sqcup \chi_2))$, where $\mathcal{T}_\mathsf{J} : \mathcal{F}_m \to \mathcal{S}$ is such that for each $\chi \in \mathcal{F}_m$, $\mathcal{T}_\mathsf{J}(\chi)$ is consistent with χ with respect to J.*

Proof idea. The basic idea is to "encode", in the time difference of two successive output spikes, the positions of all the input spikes that have occurred since the last input gap of the form $(t, t+W)$, where $\mathcal{T}(\cdot)$ is W-resettable. Such pairs of output spikes are produced once every p milliseconds, with the time difference within each pair being a function of the time difference within the previous pair and the input spikes encountered since. Intuitively, it is convenient to think of this encoding as one from which we can "reconstruct" the entire past input spike-train ensemble after the last reset gap in the input. We first describe the encoding function for the case of a single input spike-train after which we indicate how it can be generalized.

So, suppose the time difference of the successive spikes output by J lies in the interval $[0, 1)$. Define the encoding function as $\varepsilon_0 : [0,1) \times \bar{\mathcal{S}}_{(0,p]} \to [0,1)$, that takes in the old encoding and the input spikes in the past p milliseconds to produce the new encoding, which is output by J as the time difference between a new pair of spikes. The number p is chosen to be such that there are at most 8 spikes in any interval of the form $(t, t+p]$. We now describe how $\varepsilon_0(e, \mathbf{x})$ is computed, given $e \in [0,1)$ and $\mathbf{x} = \langle x^1, x^2, \ldots, x^k \rangle$, such that each spike time in $\mathbf{x}$ lies in the interval $(0, p]$. Let e have a decimal expansion[19], so that $e = 0.c_1s_1c_2s_2c_3s_3\cdots$. Accordingly, let $c = 0.c_1c_2c_3\cdots$ and $s = 0.s_1s_2s_3\cdots$. c is a real number that encodes the number of spikes in each interval of length p encountered, since the last reset. Since each interval of length p has between 0 and 8 spikes, the digit

[19] Whenever we say decimal expansion, we forbid decimal expansions with an infinite number of successive 9s. With this restriction, each real number has a unique decimal expansion.

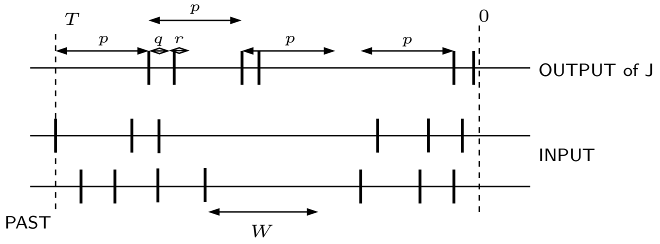


Fig. 12 This figure illustrates the operation of the intermediate neuron J. Suppose $\chi \in \mathcal{F}_m$ is an input spike-train. Let its oldest spike be T milliseconds ago. Then J produces a spike at time[20]$T - p$ and at every $T - kp$, for $k \in \mathbb{Z}^+$, unless in the previous p milliseconds to when it is to spike, there is a gap[21]of the form $(t, t + W)$. For the sake of exposition, let's call these the "clock" spikes. Now, suppose there is a gap of the form $(t, t + W)$ in the input and there is an input spike at time t, then the neuron spikes at time $t - p$ and every p milliseconds thereafter subject to the same "rules" as above. These clock spikes are followed by "encoding" spikes, which occur at least q milliseconds after the clock spike, but less than $q + r$ milliseconds after, where q is greater than the absolute refractory period α. As expected, the position of the current encoding spike is a function of the time difference between the previous encoding and clock spikes[22]and the positions of the input spikes in the p milliseconds before the current clock spike. The output of the encoding function is, in effect, appropriately scaled to "fit" in this interval of length r; the details are available in the proof.

9 is used as a "termination symbol". So, for example, suppose there have been 4 intervals of length p, since the last reset with $5, 0, 8$ and 2 spikes apiece respectively, then $c = 0.8059$ and $c' = 0.28059$, where c' is the "updated" value of c. Likewise, s is a real number that stores the positions of all input spikes encountered since the last reset. Let each spike time be of the form $x^i = 0.x^i_1 x^i_2 x^i_3 \cdots \times 10^q$, for appropriate q, whose value is fixed for a given p. Then the updated value of s is $s' = 0.x^1_1 x^2_1 \cdots x^k_1 s_1 x^1_2 x^2_2 \cdots x^k_2 s_2 \cdots$. Suppose the c' and s' obtained above were of the form $c' = 0.c'_1 c'_2 c'_3 \cdots$ and $s' = 0.s'_1 s'_2 s'_3 \cdots$, then $\varepsilon_0(e, \mathbf{x}) = 0.c'_1 s'_1 c'_2 s'_2 \cdots$. Observe that the decimal expansion constructed by $\varepsilon_0(e, \mathbf{x})$ cannot have infinitely many successive 9s, for c' has only a finite number of non-zero digits. Suppose the input were a spike-train ensemble of order m, then for each spike-train an encoding would be computed as above and in the final step, the m real numbers obtained would be interleaved together, so as to produce the encoding.

Given knowledge of the encoding function, Figure 12 briefly describes how J works. The claim then is that if two input spike-train ensembles are different at some point with no intervening "reset" gaps, then the output of J in the past U milliseconds, where $U = p + q + r$ will be different. Intuitively, this is because the difference between the latest encoding and clock spike in each case would be different, as they encode different "histories" of input spikes. The exception is if the input spike-train ensembles differed only in the past U milliseconds. In this case, the difference is communicated to O directly by χ.

Finally, we ought to remark that the above is just an informal description that glosses over several technical details contained in the full proof, which is available in Online Resource B. □

The preceding two propositions thus imply Lemma 7 which together with Lemma 6 implies Theorem 3.

Lemma 7. *If $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$ is causal, time-invariant and resettable, then it can be effected by a feedforward network of depth two.*

Theorem 3. *If $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$ can be effected by a feedforward network, then it can be effected by a feedforward network of depth two.*

Corollary 2. *The set of all feedforward networks is not more complex than the set of feedforward networks of depth equal to two.*

Incidentally, Lemma 6 and 7 also lead to a full characterization of the class of transformations effected by all feedforward networks equipped with neurons obeying the abstract model of Section 3. This is formalized in the next theorem.

Theorem 4. *A transformation $\mathcal{T} : \mathcal{F}_m \rightarrow \mathcal{S}$ can be effected by a feedforward network if and only if it is causal, time-invariant and resettable.*

Directions for further constraining the present model

The results of this section imply that we need to add new properties to further constrain our model neurons, in order for complexity results involving feedforward networks of depth two to be manifested. There are a number of directions that one could take. One is that spike-times in the present model are real numbers. When stochastic variability in neurons is taken into account, this assumption is no longer true. Also, we did not assume that the membrane potential changes smoothly with time, which would be a reasonable assumption to add. And, finally, an assumption consistent with Dale's principle, that each neuron has either an excitatory effect on all its postsynaptic neurons or an inhibitory effect might also help in this direction.

10 Discussion

There has been some debate about how useful data from the connectome projects might be in advancing a mechanistic understanding of computation occurring in the circuits of the brain. One of the main type of arguments that has

[20] i.e. p milliseconds after time instant T.

[21] We set $W > p$ to force a spike at $T - p$.

[22] unless the present clock spike is the first after a reset gap in the input.

been made against their utility is that, since these projects only[23] seek to ascertain the wiring diagram, without giving us detailed physiological information, it is not clear what we might learn from this data alone, especially for networks whose high-level function is not known. While it is acknowledged that network architecture places constraints on what a network can compute (Kleinfeld et al, 2011; Denk et al, 2012), the nature and scope of these constraints have remained poorly understood. Our goal with this work was in asking, on one hand, if we can deduce non-trivial examples of computations that a network *could not* be doing, given just the knowledge of its architecture and assuming that the neurons obey some elementary properties. On the other hand, we asked if there are fundamental limits to what can be said, given just this information. We examined this question for the case of feedforward networks equipped with neurons that obeyed a deterministic spiking neuron model. We first set the stage by creating a mathematical framework in which this question could be precisely posed. Crucially, we needed to make precise what computation exactly meant in this context. This took a fair bit of work and led us to the view of feedforward networks as spike-train to spike-train transformations under biologically-relevant spiking regimes. After setting up necessary definitions, we then showed some examples of transformations that networks of specific architectures *cannot* effect, that other networks can. First of all, we showed[24] that there exist spike-train to spike-train transformations that no feedforward network could effect. Next, we showed a transformation that no single neuron could effect but a network consisting of two neurons could. After this, we proved a result which shows that a class of architectures that share a certain structural property also share their inability to effect a particular class of transformations. Notably, while this class of architectures has networks with arbitrarily many neurons, we showed a class of networks with just two neurons which could effect this class of transformations. This suggests that network structure alone may impose crucial constraints on computational ability. Finally, we demonstrated that the small number of properties assumed for our model neurons can only take us so far. We proved that without making further assumptions about our model neurons, we couldn't discern such examples for the set of all feedforward networks of depth two.

While there is more to neuronal networks than just their wiring diagram, what our theory suggests is that the wiring diagram could impose crucial constraints on the computational ability of networks, in some cases. On the other hand, there seem to be classes of networks for which a more elaborate knowledge of single neuron properties may be necessary, before we can determine restrictions on their computational ability. While technical issues in electron microscopy (Denk et al, 2012) have so far stood in the way of mapping, for example, distributions of ion-channels and neurotransmitter and neuromodulator receptors in neurons, it is conceivable that such hurdles may be overcome in future. If successful, these or other advances in conjunction with the wiring diagram could provide useful information to help us tease out pertinent constraints on the computational capabilities of these networks.

In this work, as a first step, we have aimed to demonstrate specific *examples* of computations that a network cannot accomplish, given its architecture. The more ambitious goal would be the ability to have an exact characterization of the set of *all* computations that a given neural circuit cannot perform, given knowledge of its architecture, to the extent that a given incomplete knowledge of the physiological properties of its neurons will allow. This is not necessarily a goal that is out of reach. Even in the present work, we have obtained such an exact characterization[25] of the set of all computations that the set of feedforward networks cannot accomplish, given the set of properties that our model neurons are presently assumed to obey. Therefore, in principle, there seems to be no reason why we may not be able to do likewise for specific network architectures.

Acknowledgements This work was supported, in part, by a National Science Foundation grant (NSF IIS-0902230) to A.B.

References

Banerjee A (2001) On the phase-space dynamics of systems of spiking neurons. I: Model and experiments. Neural Comp 13(1):161–193

Bi Gq, Poo Mm (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. The Journal of Neuroscience 18(24):10,464–10,472

Bock DD, Lee WCA, Kerlin AM, Andermann ML, Hood G, Wetzel AW, Yurgenson S, Soucy ER, Kim HS, Reid RC (2011) Network anatomy and in vivo physiology of visual cortical neurons. Nature 471(7337):177–182

Briggman KL, Helmstaedter M, Denk W (2011) Wiring specificity in the direction-selectivity circuit of the retina. Nature 471(7337):183–188

Chklovskii DB, Vitaladevuni S, Scheffer LK (2010) Semi-automated reconstruction of neural circuits using electron microscopy. Current opinion in neurobiology 20(5):667–675

[23] This in itself is a formidable problem and one that is taking heroic effort.

[24] See Figure 7(a) and the second paragraph of Section 8.

[25] This characterization is a consequence of Theorem 4. In particular, it is the set of all transformations that are *not* causal, time-invariant or resettable.

Denk W, Horstmann H (2004) Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS Biology 2(11):e329

Denk W, Briggman KL, Helmstaedter M (2012) Structural neurobiology: missing link to a mechanistic understanding of neural computation. Nature Reviews Neuroscience 13(5):351–358

Hayworth K, Kasthuri N, Schalek R, Lichtman J (2006) Automating the collection of ultrathin serial sections for large volume tem reconstructions. Microsc Microanal 12(Suppl 2):86–87

Helmstaedter M, Briggman KL, Denk W (2011) High-accuracy neurite reconstruction for high-throughput neuroanatomy. Nature neuroscience 14(8):1081–1088

Helmstaedter M, Briggman KL, Turaga SC, Jain V, Seung HS, Denk W (2013) Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 500(7461):168–174

Kleinfeld D, Bharioke A, Blinder P, Bock DD, Briggman KL, Chklovskii DB, Denk W, Helmstaedter M, Kaufhold JP, Lee WCA, et al (2011) Large-scale automated histology in the pursuit of connectomes. The Journal of Neuroscience 31(45):16,125–16,138

Knott G, Marchman H, Wall D, Lich B (2008) Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. The Journal of Neuroscience 28(12):2959–2964

Markram H, Lübke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. Science 275(5297):213–215

Mikula S, Binding J, Denk W (2012) Staining and embedding the whole mouse brain for electron microscopy. Nature methods 9(12):1198–1201

Mishchenko Y, Hu T, Spacek J, Mendenhall J, Harris KM, Chklovskii DB (2010) Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. Neuron 67(6):1009–1020

Nirenberg S, Carcieri S, Jacobs A, Latham P (2001) Retinal ganglion cells act largely as independent encoders. Nature 411(6838):698–701

Reid RC (2012) From functional architecture to functional connectomics. Neuron 75(2):209–217

Rieke F, Warland D, van Steveninck R, Bialek W (1997) Spikes: exploring the neural code. MIT Press, Cambridge, MA

Seung HS (2011) Towards functional connectomics. Nature 471(7337):170–172

Shepherd G (2004) The synaptic organization of the brain. Oxford University Press, New York, NY

Strehler B, Lestienne R (1986) Evidence on precise time-coded symbols and memory of patterns in monkey cortical neuronal spike trains. Proc Nat Acad Sci USA 83(24):9812

Takemura Sy, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P, et al (2013) A visual motion detection circuit suggested by drosophila connectomics. Nature 500(7461):175–181

Turaga SC, Murray JF, Jain V, Roth F, Helmstaedter M, Briggman K, Denk W, Seung HS (2010) Convolutional networks can learn to generate affinity graphs for image segmentation. Neural Computation 22(2):511–538