bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Parallel and recurrent cascade models as a unifying force for understanding sub-cellular computation

Emerson F. Harkin[a,*], Peter R. Shen[b,*], Anish Goel[c], Blake A. Richards[d,e,f,g,**], Richard Naud[a,h,**]

[a] *uOttawa Brain and Mind Institute, Centre for Neural Dynamics, Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada*
[b] *Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada*
[c] *Lisgar Collegiate Institute, Ottawa, ON, Canada*
[d] *Mila, Montréal, QC, Canada*
[e] *Montreal Neurological Institute, Montréal, QC, Canada*
[f] *Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada*
[g] *School of Computer Science, McGill University, Montréal, QC, Canada*
[h] *Department of Physics, University of Ottawa, Ottawa, ON, Canada*

Abstract

Neurons are very complicated computational devices, incorporating numerous non-linear processes, particularly in their dendrites. Biophysical models capture these processes directly by explicitly modelling physiological variables, such as ion channels, current flow, membrane capacitance, etc. However, another option for capturing the complexities of real neural computation is to use cascade models, which treat individual neurons as a cascade of linear and non-linear operations, akin to a multi-layer artificial neural network. Recent research has shown that cascade models can capture single-cell computation well, but there are still a number of sub-cellular, regenerative dendritic phenomena that they cannot capture, such as the interaction between sodium, calcium, and NMDA spikes in different compartments. Here, we propose that it is possible to capture these additional phenomena using parallel, recurrent cascade models, wherein an individual neuron is modelled

*These authors contributed equally.
**Co-corresponding authors, order decided by coin toss.
Email addresses: blake.richards@mila.quebec (Blake A. Richards), rnaud@uottawa.ca (Richard Naud)

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as a cascade of parallel linear and non-linear operations that can be connected recurrently, akin to a multi-layer, recurrent, artificial neural network. Given their tractable mathematical structure, we show that neuron models expressed in terms of parallel recurrent cascades can themselves be integrated into multi-layered artificial neural networks and trained to perform complex tasks. We go on to discuss potential implications and uses of these models for artificial intelligence. Overall, we argue that parallel, recurrent cascade models provide an important, unifying tool for capturing single-cell computation and exploring the algorithmic implications of physiological phenomena.

Keywords: Cascade models, Single-cell computation, Dendritic non-linearities, Artificial neural networks

---

Introduction

One of the greatest success stories in modern neuroscience is the development of biophysical models of single-cell computation. Though there are still many mysteries to be explored, and new discoveries are still being made, our core understanding of the biophysics of excitable membranes as described by circuit equivalence models, cable theory, and Hodgkin & Huxley-style models has stood the test of time and can reasonably be considered as an accepted theory in neuroscience (Brunel et al., 2014, Herz et al., 2006, Hodgkin and Huxley, 1952, Koch, 2004, McKenna et al., 2014, Rall, 1960). This has provided the foundation for countless computational studies on single-cell computation using detailed models that explicitly incorporate physiological variables including membrane capacitance, ion channels, reversal potentials, etc. Such models have proven very useful for understanding a variety of phenomena, particularly dendritic computation (Poirazi and Papoutsi, 2020, Mainen and Sejnowski, 1996, Schaefer et al., 2003, Vetter et al., 2001, Shai et al., 2015, Psarrou et al., 2014, Gidon et al., 2020, Krichmar et al., 2002, Cook and Johnston, 1997, Gasparini et al., 2004, Ariav et al., 2003, Eyal et al., 2014, Deitcher et al., 2017, Papoutsi et al., 2017). Without these models we would understand much less than we do about how dendrites contribute to computation in neural circuits.

However, due to their complexity, biophysical models are very difficult to link to algorithmic models of neural computation. To some extent, this is part of the dilemma we always face in science, i.e., "How detailed should our models be in order to achieve our scientific goals?" (Poirazi and Pa-

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

poutsi, 2020, Richards et al., 2019). But, one thing that we can say is that it would be beneficial for understanding the functional importance of dendritic computation if we possessed models of intermediate complexity that could capture many of the phenomena of single-neuron computation while still being sufficiently mathematically tractable to use for explaining complex animal behaviour. Moreover, if we could develop such intermediate models we would be better placed to use insights on dendritic computation to inform the development of novel machine learning approaches (Richards et al., 2019, Sinz et al., 2019).

To this end, previous work has explored the use of "cascade models" to capture dendritic computation (Poirazi et al., 2003, Tzilivaki et al., 2019, Kalmbach et al., 2017, Ujfalussy et al., 2018) (but see (Francioni and Harnett, 2021). Cascade models use a cascade of linear and non-linear operations, which are effectively tree-structured, feedforward, multi-layer artificial neural networks (ANNs) (Poirazi and Papoutsi, 2020). Research has shown that these models can capture more variance in single-cell activity than standard point neuron models (which consist of a single linear operation and non-linear activation function) (Poirazi et al., 2003, Tzilivaki et al., 2019, Ujfalussy et al., 2018). Thus, cascade models treat individual neurons as deep feedforward ANNs in order to capture single-cell computation with a mathematically tractable model (Figure 1A). However, such models still have not been compared to many facets of dendritic computation, including calcium spikes and N-methyl-D-aspartate (NMDA) receptor mediated plateaus.

Here, we show that it is possible to capture these phenomena using **parallel and recurrent cascade models** (PRC models), i.e. models that use a set of parallel cascades of linear and non-linear operations that are recurrently connected to one another. This is equivalent to treating individual neurons as multi-layer, recurrent ANNs (Figure 1B). We show that these PRC models can successfully reproduce a number of experimentally observed regenerative phenomena in dendrites, all while being mathematically tractable. Moreover, because these models are fully differentiable, we show that they can easily be incorporated into machine learning approaches that utilize gradient descent for model optimization (Richards et al., 2019). This opens the door to exploring the possibility that dendrites and dendritic computation may provide important inductive biases for brains that could be mimicked by artificial intelligence (AI) (Richards et al., 2019, Sinz et al., 2019). Thus, we conclude this paper by providing some speculation as to the possible advantages for AI of dendritic PRC models. Altogether, this work

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

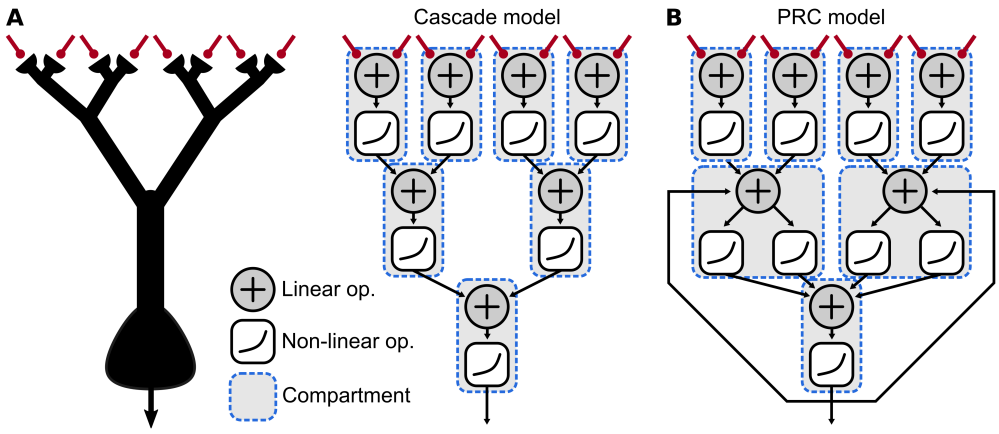

Figure 1: Illustration of cascade models and PRC models: **A** Dendritic computation can be modelled using a cascade of linear and non-linear operations (Poirazi et al., 2003, Tzilivaki et al., 2019, Ujfalussy et al., 2018). **B** Cascade models can be extended with the use of parallel pathways and recurrence in the operations, which can enable the modeling of more complicated regenerative phenomena.

helps to lay the ground for better integration between our well-developed understanding of the biophysics of neural computation and our ever increasing understanding of algorithms for complex behaviour.

The use of recurrent cascade models to capture single-cell computation

Models made of a cascade of linear-nonlinear (LNL) operations have had a long history in systems neuroscience, where such models were conceptually implied by early work on retinal ganglion cells (Kuffler, 1953) and cortical cells (Hubel and Wiesel, 1968). These models consist of a linear filter of the stimulus followed by a nonlinear readout to generate predictions of a non-negative observable. Important refinements to improve the accuracy of these models were the inclusion of spike-triggered adaptation (Pillow et al., 2005, Truccolo et al., 2005), the composition of linear-nonlinear (LNL) operations in a cascade (Vintch et al., 2015, McFarland et al., 2013), and the addition of recurrent interactions between linear-nonlinear subunits (Pillow et al., 2008, Truccolo et al., 2011). Together, these various extensions are described by an LNL subunit with multiple possible types of interconnection motifs. The LNL subunit is composed of a linear-nonlinear operation with feedback from

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the output of the nonlinearity to capture spike-triggered adaptation (Fig. 2A; Methods 1.1). Wiring between subunits can create multiple types of motifs including a strictly feedforward cascade, a common input to two units having possibly different filters or nonlinearities (parallel feature processing, Fig. 2B), or recurrent interconnections (Fig. 2C). In most systems neuroscience applications, the output of the nonlinearity is the Poisson intensity, used to capture the stochastic spike-timing responses of real neurons. Our approach with PRC models here is slightly different, as we will consider the output of the nonlinearity to represent deterministic voltage excursions. Also, in most systems neuroscience applications, the filters and nonlinearities may arise from a large number of possible – yet undefined – mechanisms. These are typically thought of as interactions within and between cells, but may also include dendritic computations (Taylor et al., 2000). In order to better understand the cellular mechanisms underlying such information processing, we focus on the application of a deterministic LNL framework within a single cell.

Somatic spikes

Since the pioneering work of Richard Stein (Stein, 1965), the leaky integrate-and-fire (LIF) model has become the workhorse of interrogations of information processing of large numbers of interconnected neurons. In itself, the LIF can be seen as a special case of the LNL subunit. When a depolarizing current is injected into an LIF model, it first passes through the membrane filter and produces a voltage (leaky integration; a linear operation) which is eventually reset to a lower value if it reaches a threshold (firing; a nonlinear operation). This leaky integrate-and-fire behaviour can be captured by an LNL subunit with a filter that corresponds to the membrane filter of an LIF model, together with a Heaviside nonlinearity that is triggered exactly at spike threshold and a Dirac delta-shaped adaptation filter which resets the voltage to a lower value. Such deterministic LNL models can capture both the time-dependent membrane voltage response and the spike timing responses to complex inputs (Jolivet et al., 2006). Adding multiple time-scales to the kernel for spike-triggered adaptation makes these models highly predictive of the responses of a variety of cell types (Jolivet et al., 2008, Mensi et al., 2012, Pozzorini et al., 2013, Kobayashi et al., 2009, Teeter et al., 2018). Furthermore, considering smooth nonlinearities and surrogate gradients can ensure that the LNL unit remains differentiable (Neftci et al., 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

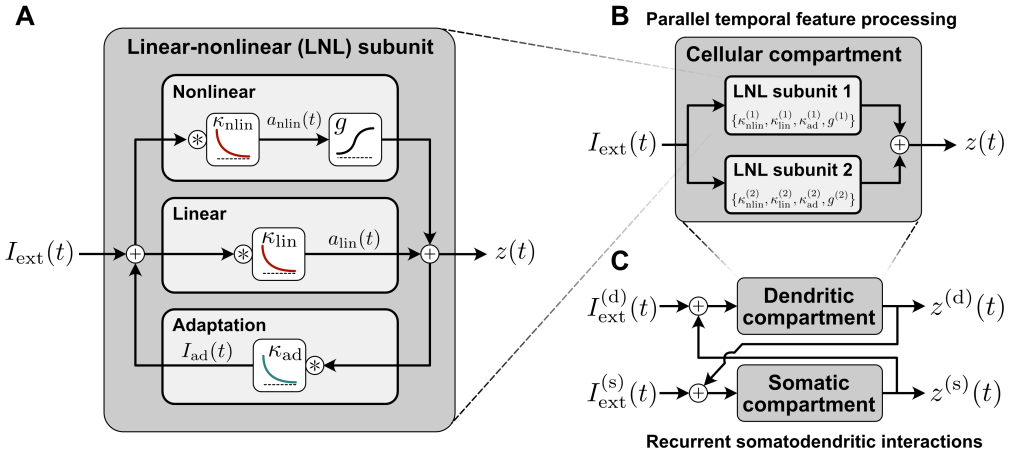

Figure 2: Linear-nonlinear operations and interconnection motifs. **A** The linear-nonlinear (LNL) subunit. This architecture combines linear filtering of an input $I_{\mathrm{ext}}(t)$ (middle) with a nonlinear readout (top, g) and feedback (bottom, κ_{ad}), which together flexibly capture the passive filtering effects of neuronal membranes, the nonlinear effects of voltage-dependent ionic conductances, and adaptation. The contributions of each of these effects to the output $z(t)$, which may loosely represent neuronal voltage or a spiketrain, can be tuned via the parameters of the filters κ_{nlin}, κ_{lin}, and κ_{ad}, and the choice of nonlinear function $g(\cdot)$. **B** Multiple LNL subunits arranged in parallel can model the effects of multiple ionic conductances in a single cellular compartment. **C** Recurrent connections between model compartments, each composed of one or more LNL subunits, capture interactions between cellular compartments.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

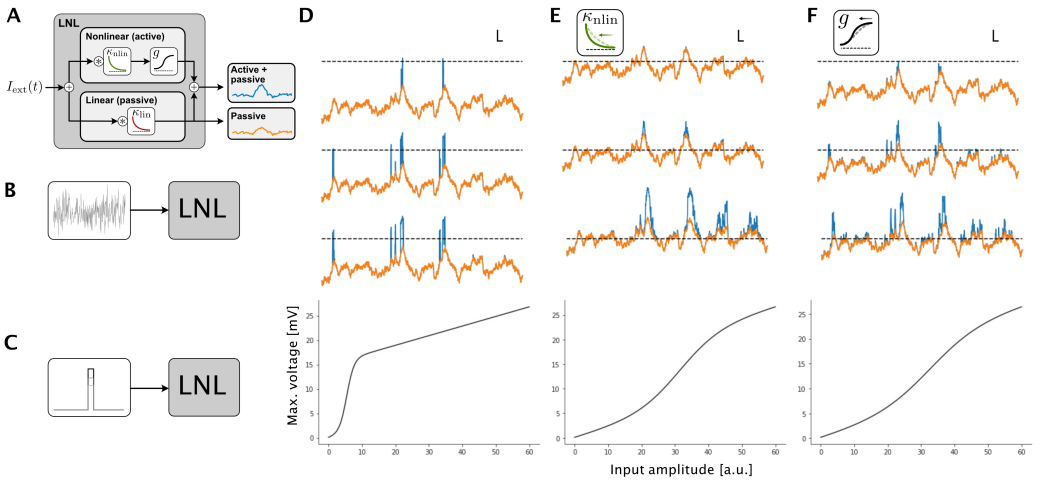

Figure 3: Effects of changing parameters in the linear-nonlinear model. **A** Schema of the LNL model. An external input undergoes two parallel processing steps: a linear (passive) filter and a linear filter (possibly distinct from passive) followed by a nonlinear readout. The nonlinear processing step (top) produces a nonlinear component (blue) which is added to a passive component (orange). As input to the model, we consider both **B** a noisy time-dependent signal representing a bombardment of asynchronous synaptic currents, and **C** a short pulse. **D, top 3 traces** Response to noisy inputs having three different baselines (lower baseline is topmost trace,dashed line represents the activation threshold of the nonlinearity). The nonlinear (blue) is added to the passive component (orange). **D, bottom** Maximum amplitude of response as a function of pulse amplitude (bottom). **E** Same as D but using a model with an increased timescale of the linear filter. **F** Same as D but with a model having a decreased sensitivity (steepness) of the nonlinearity. Scale bars correspond to 10 ms and 5 mV.

In addition it is worth noting that since LNL subunits are a special case of PRC models, and LIF models are a special case of LNL models, LIF models are, in fact, a restricted class of PRC models. As such, they can capture some of the phenomena that the broader class of PRC models are capable of capturing (Weber and Pillow, 2017, Gerstner et al., 2014). Part of our contribution here is to illustrate some of the more complicated sub-cellular phenomena that cannot be recapitulated with a pure LIF model, such as dendritic non-linearities and coincidence detection (see sections 2.3-5, below).

LNL models for dendritic spikes

To circumscribe a systems-level function for dendritic computation, many studies have focused on the role of intrinsic nonlinear dendritic operations—first using models of dendritic trees in a stationary state (Mel, 1993, Poirazi

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and Mel, 2001, Poirazi et al., 2003), then using models capturing the dynamics of dendrites and their intrinsic nonlinear properties (Legenstein and Maass, 2011, Kalmbach et al., 2017, Ujfalussy et al., 2018). These contributions are examples of what we call PRC models because they involve a cascade of nonlinearities, but they leave out the recurrent motifs. Also, most have not considered the parallel processing introduced by Ujfalussy et al. (2018) (Ujfalussy et al., 2018). Recurrence was, however, part of other efforts focusing on simplified models of the interaction between somata and dendrites (Pinsky and Rinzel, 1994, Urbanczik and Senn, 2014, Naud et al., 2014, 2017). Thus our goal in this section is to unify these complementary perspectives by presenting data showing that PRC models can reproduce qualitative features of sub-cellular physiological phenomena, such as dendritic spikes, coincidence detection, etc.

In comparison to the action potentials generated in the proximity of the cell body, the spikes observed in dendrites display less stereotypical amplitudes and durations (Golding and Spruston, 1998, Larkum and Zhu, 2002, Smith et al., 2013, Schiller et al., 2000). These features pose a problem for the LIF framework, but they are captured naturally by the LNL framework. In Figure 3, we illustrated the response of a LNL subunit to noisy inputs and to short pulses. For a fast filter preceding a sharp nonlinearity, the model produced short spikes on top of a noisy voltage trace. The short spikes arise when the low-pass filtered input reaches the activation threshold of the nonlinear readout. Sometimes, the fluctuation is only able to activate a fraction of the nonlinearity, which produces spikes of variable amplitudes. Increasing the mean of the input makes those spikes more frequent as there are more chances that the random fluctuations reach the activation threshold of the nonlinearity. Thus in the configuration where a sharp sigmoidal nonlinearity is fed by a linear filter that is considerably faster than the membrane filter, we observe variable amplitude spikes akin to dendritic sodium spikes. We note that a very similar model architecture was able to be reproduce with great precision the response to noisy dendritic inputs in the presence of dendritic sodium spikes (Kalmbach et al., 2017).

Next, we examined the effects of changing the parameters in the LNL subunit (Fig. 3E). We began by increasing the timescale of the filter preceding the nonlinearity. This reduced the number of suprathreshold fluctuations, and when a fluctuation in the low-pass filtered input did cross the activation threshold it tended to stay for a longer time period. This produced less frequent but longer spikes, akin to calcium spikes (Larkum et al., 1999, Larkum

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and Zhu, 2002, Magee and Carruth, 1999, Xu et al., 2012) or, with an even longer timescale, NMDA spikes (Schiller et al., 2000). As a consequence of changing the filter, the aspect of the nonlinearity that can be observed when presenting a pulse input is altered, and appears more graded. When, instead of changing the timescale of the linear filter, we only changed the gain of the nonlinearity (Fig. 3F), then the spikes had a more variable amplitude and duration. Changing the timescale of the filter and the steepness of the nonlinearity had a similar effect on the maximum voltage observed in response to a pulse input. This can be explained by the fact that in response to a short pulse, the voltage increases linearly with a slope given by the timescale of the filter (Gerstner et al., 2014) and thus acts as a multiplicative factor identical to the steepness of the nonlinearity. In these simulations, we have not included an adaptation-like recurrence, although the formalism can include this mechanism. Thus, altogether, by changing the parameters of the LNL model, we can simulate some basic electrophysiological features of various types of dendritic spikes.

Dendritic sodium spikes

To test whether a PRC model can capture other features observed in electrophysiological recordings, we focused on experimental findings reported by Golding and Spruston (1998) (Golding and Spruston, 1998) pertaining to dendritic sodium spikes. In one of the experiments reported (Fig. 4), recordings were made simultaneously in a proximal dendrite and in the soma. A variable-amplitude synaptic-like stimulus was injected in the dendrite. The recordings showed that in one dendrite, low input amplitudes initiated a spike in the soma which produced a back-propagating action potential in the dendrite and at high amplitudes initiated a spike in the dendrite <u>before</u> triggering an action potential in the soma. In another dendritic recording, a low amplitude stimulus initiated a dendritic spike unaccompanied by an action potential at the soma, and only a large input produced an action potential at the soma. We found that we can reproduce these phenomena by changing the parameters of two LNL subunits wired in a recurrent fashion (Fig. 4A and C). To capture how different recordings initiated spikes preferentially in the dendrite or the soma, we varied the relative threshold of activation of the dendritic and somatic nonlinearities.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

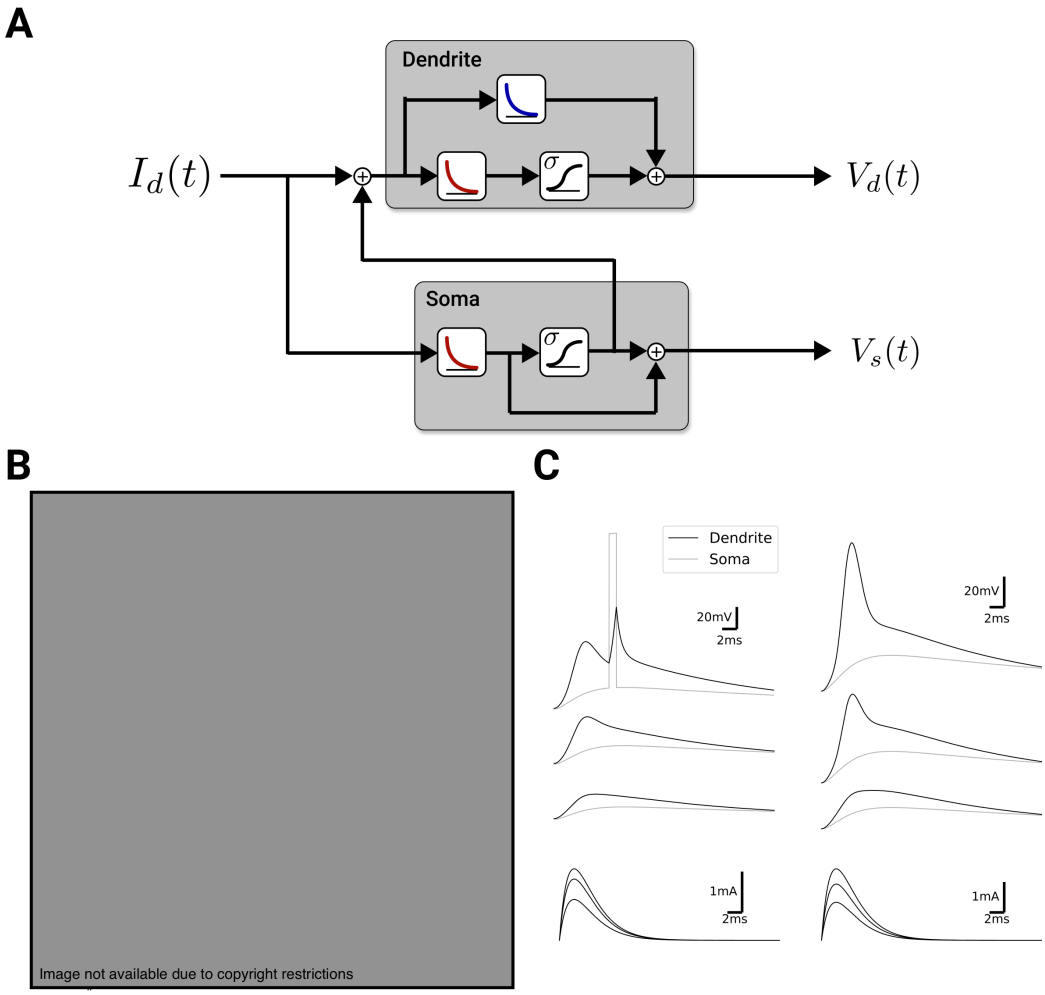

Figure 4: A recurrent motif of linear-nonlinear models for the dendritic sodium spikes. **A** Schematic of the model: A dendritic current (I_d) impinges on two LNL subunits with recurrent interactions, one corresponding to a dendrite, another corresponding to the soma. When the somatic compartment reaches the threshold of its nonlinearity, a spike in the form of a square pulse is added to the dendritic current. **B** Experimental data showing injection of synaptic-like pulses of varying amplitudes in the dendrite (topmost traces have lowest input), where the voltage in the dendrite (thick trace) and soma (thin trace) are shown. Two exemplars are shown in different columns. Reproduced from Golding et al. (1998) Fig. 1 (Golding and Spruston, 1998). **C** Model responses showing three amplitude levels of a synaptic-like input in the model shown in A. To reproduce the exemplars in B the model in the right column has a lower threshold for the dendritic nonlinearity and a higher threshold for the somatic nonlinearity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

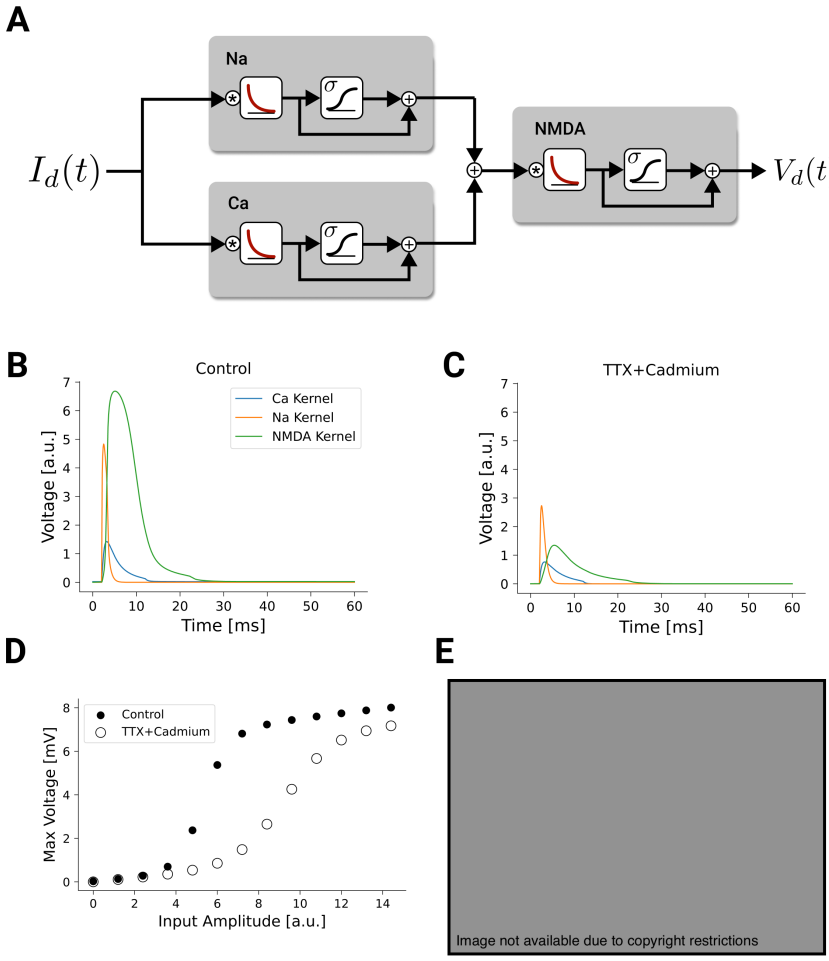

Figure 5: Linear nonlinear model of a NMDA spike as a combination of cascade and parallel processing. **A** Schematic of the model: current impinging on the dendrite is passed through two LNL operations in parallel before feeding into another LNL operation to produce an NMDA spike. **B** Response of the model to a supra-threshold pulse input showing the calcium (blue) sodium (orange) and NMDA (green). **C** The response to the model with the same pulse input as in B but with the nonlinear component of sodium and calcium set to zero, simulating the application of TTX and Cadmium. **D** Maximum voltage as a function of the amplitude of the input pulse for the model in B (full circles) and the model in C (open circles). **E** Experimental recordings of peak membrane potential as a function of stimulation power in control (full circles) and the presence of calcium and sodium ion channel blockers (TTX and cadmium, open circles). Figure reproduced from Schiller et al. (2000) Fig. 3c (Schiller et al., 2000).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dendritic NMDA spikes

Next we considered the electrophysiological recordings of NMDA spikes reported in Schiller et al. (2000) (Schiller et al., 2000) (Fig. 5). We focused on the threshold input-pulse amplitude to trigger an NMDA spike which was lowered by the addition of blockers of sodium and calcium ion channels (TTX and cadmium). This observation suggested that the nonlinear effects of sodium and calcium ion channels participated in the initiation of the NMDA spikes. Since calcium and sodium ion channels are characterized by distinct timescales, we considered a parallel connectivity motif shown in Fig. 5A. We chose the filter of the sodium LNL to be fast (5 ms), the filter of the calcium filter to be slower (40 ms), and the filter of the NMDA LNL to be even slower (80 ms). Simulating the response of this model to pulse inputs produced a long depolarization that was clearly initiated by contributions from sodium and calcium (Fig. 5B). To simulate the blockade of these mechanisms by TTX and cadmium, we reduced the amplitudes of their corresponding nonlinearities to zero, which prevented the occurrence of spikes for a range of input amplitudes (Fig. 5C). The effect on the initiation threshold of removing the nonlinearity in the PRC model mimicked that of pharmacological manpulations (Fig. 5D-E).

Dendritic calcium spikes

We then considered how bidirectional interactions between somatic spiking and calcium spikes can be captured by a PRC model. The tuft dendrites of cortical pyramidal cells express a high density of voltage-gated calcium channels which produce dendritic plateau potentials when sufficiently strong inputs are injected into the soma and tuft dendrites simultaneously (Larkum et al., 1999, Magee and Carruth, 1999). These dendritic plateau potentials mediate burst firing at the soma, producing coincidence detection and modulating the gain of somatic responses to peri-somatic input. To capture these effects in our PRC framework, we created a model with two recurrently-connected compartments, corresponding to the soma and apical tuft dendrites (Fig. 6A). Appropriately tuned filtering and nonlinear operations in each compartment (see Methods 1.4) caused the somatic compartment to emit single spikes when input was delivered to the soma alone and intermittent bursts when input was delivered to both compartments simultaneously (Fig. 6B). Inputs delivered to either compartment alone evoked small responses in the dendritic compartment, while simultaneous inputs to both compartments evoked burst-associated plateau potentials in the simulated

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dendrite. In cortical pyramidal neurons, dendritic inputs produce somatic bursts most potently when they are delivered just before or at the same time as somatic input, creating an asymmetric coincidence-detection effect (Larkum et al., 1999). Injecting a synaptic-like pulse into the dendritic compartment of our model evoked a dendritic plateau potential and somatic burst only when it preceded or arrived at the same time as a somatic input pulse (Fig. 6C), recapitulating the coincidence-detection effect observed in pyramidal neurons (Larkum et al., 1999). Dendritic input to our model also modulates somatic gain by increasing the number of spikes evoked by a given input (Fig. 6D), consistent with an effect of dendritic input on gain observed experimentally (Larkum et al., 2004). Together, these simulations add to previous efforts (Naud et al., 2014) in showing that the PRC framework is able to capture features of dendritic excitability and somatodendritic interactions.

Potential applications of recurrent cascade models for learning theory

Aside from the additional capabilities to capture biological phenomena in dendrites that we have illustrated here, PRC models may have relevance for machine learning applications. Notably, thanks to the use of differentiable computational graphs (see (Zenke and Ganguli, 2018) for an approach to making our somatic units differentiable), a PRC model can be incorporated into any artificial neural network model and trained with gradient descent (Ujfalussy et al., 2018, Jones and Kording, 2021). As such, PRC models open the door to investigating whether sub-cellular dendritic computations have any potential utility for improving machine learning applications. The answer to this question will depend, in large part, on whether dendritic computations can provide important inductive biases for an artificial neural network (Bird and Cuntz, 2020, Jones and Kording, 2021).

Networks of PRC models can be trained using gradient descent

To illustrate that complex PRC models can be incorporated into multilayer artificial neural networks and while remaining trainable via gradient descent, we next trained an artificial neural network containing a single hidden layer made of four PRC neurons to solve a simple binary classification problem: memorizing labels associated with random patterns of synaptic input (Fig. 7A). Before training, our network correctly predicted the label

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

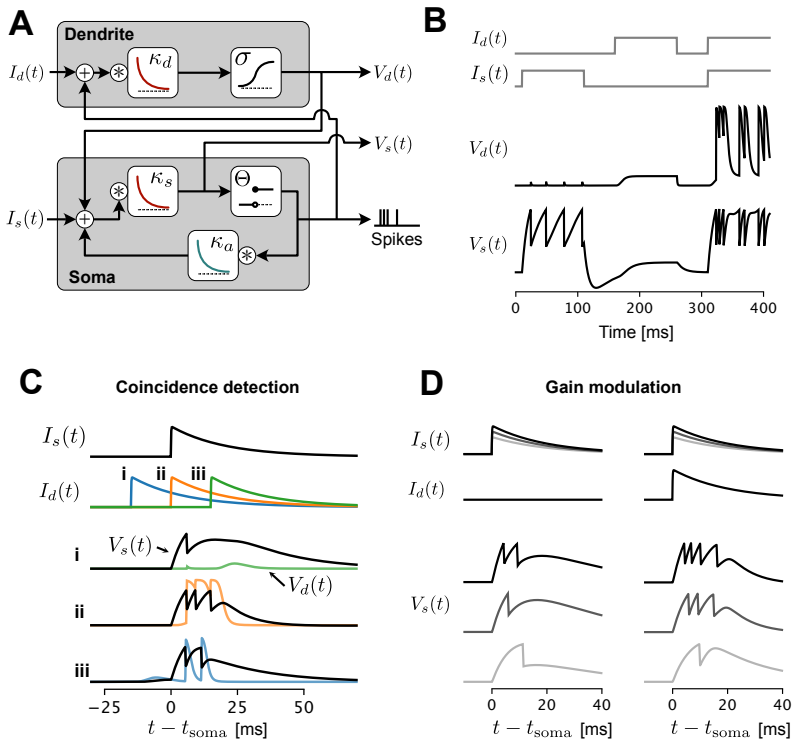

Figure 6: A recurrent cascade model of the interaction between the back-propagating action potential and calcium spikes. **A** Schematic of the model: external current impinging on the dendrite (I_d) and soma (I_s) each engage a LNL operation. The output of the nonlinearities is combined to the input current of the other unit, forming a recurrent interaction. **B** Step current injections in the soma alone, the dendrite alone and then in both compartments simultaneously produces regular bursting only for the conjunction of inputs. **C** A strong synaptic-like current pulse is injected in the both the soma (I_s) and the dendrite (I_d). Three simulations are shown for three relative timings of the dendritic input (i blue, ii orange, iii green). Responses for the somatic (V_s, black traces) and dendritic (V_d, color traces) compartments are overlaid for each condition. Somatic spikes are denoted by a clear reset but the full depolarization is not shown. A burst of action potentials arise from the near-coincident condition (ii, orange) recapitulating experimental observations in pyramidals of the cortex (Larkum et al., 1999). **D** Responses to increasing amplitudes of synaptic-like input to soma in the absence of dendritic input (left) and in the presence of a concomitant input in the dendrite (right), a simulation of the gain modulation property of dendritic input reported experimentally (Larkum et al., 2004).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

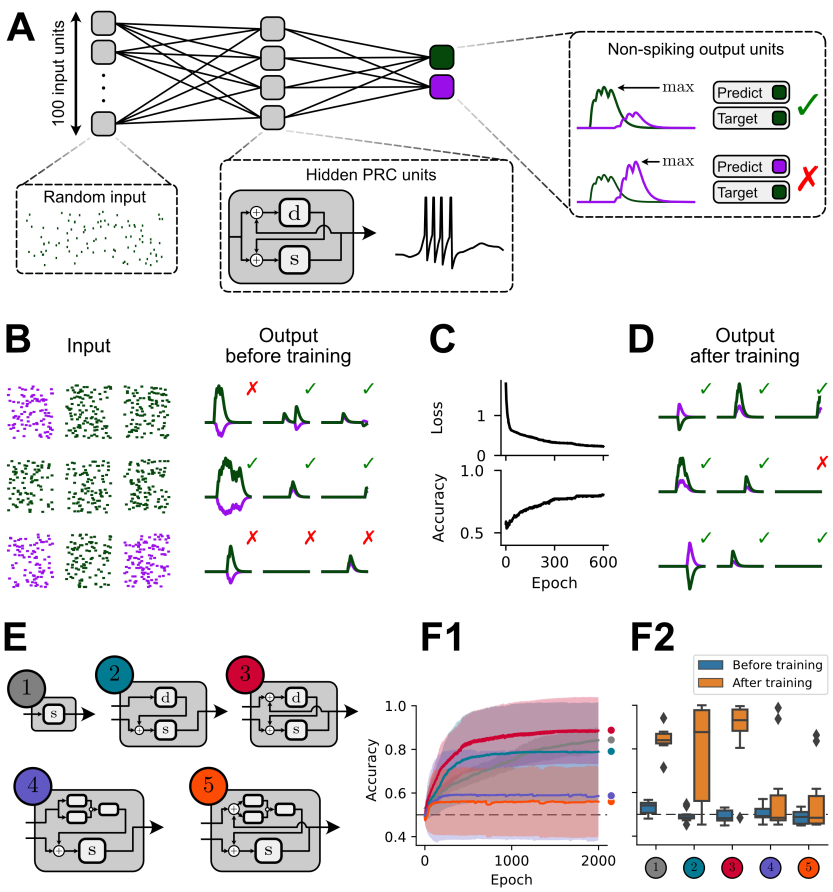

Figure 7: PRC neurons in a network can be trained to memorize patterns of synaptic input. **A** Network architecture with PRC units in the hidden layer. 100 input units send synaptic input to a hidden layer composed of four PRC neurons with recurrently-connected somatic and dendritic subunits (as shown in inset and Fig. 6A), which in turn send synaptic input to an output layer consisting of two non-spiking single-compartment neurons. Synaptic input patterns are classified as "green" or "purple" based on the maximum voltage attained in the output layer. **B** Randomly-generated patterns of synaptic inputs (left) are incorrectly classified approximately 50 % of the time by a PRC network with randomly-initialized synaptic weights. **C**, **D** Training synaptic weights using back-propagation through time allows the network to memorize labels associated with specific input patterns, resulting in improved predictions. The mean negative log likelihood loss is shown (top). A surrogate gradient was used for the Heaviside nonlinearity in the somatic compartment of hidden PRC neurons (see Methods) in order to train the input weights. **E** Different architectures of the PRC unit used in the hidden layer: ① one-compartment neuron, ② two-compartment neuron with no recurrent connection from the soma to dendrite, ③ two-compartment neuron with recurrent connections between soma and dendrite (as in 2C and 6), ④ two-compartment neuron with parallel processing in the dendritic compartment (as in 5), and ⑤ two-compartment neuron with both parallel processing and recurrent somatodendritic connections. **F1** Mean ± SD accuracy during training for $N = 10$ randomly-initialized networks of each type. Dashed lines indicate chance-level accuracy. **F2** Training set accuracy before and after training is shown for the different models.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

("green" or "purple") that had been randomly assigned to each randomly-generated pattern of synaptic input ∼50 % of the time, no better than chance. Adjusting the synaptic weights in the direction of decreasing error using back-propagation through time (with a surrogate gradient for the somatic spiking function; see Methods) gradually improved the prediction accuracy to well above the chance level (>75 %, Fig. 7C, D), indicating that the network had successfully learned a mapping from synaptic input patterns to class labels. We repeated the training procedure with different types of hidden units (Fig. 7E), ranging from an integrate-and-fire unit without a dendrite (model ①) to the full PRC model (model ⑤), and including recurrently-connected compartments without paralel processing (as in Fig. 6). For the models with a dendrite (②–⑤ in Fig. 7), two independent sets of weights were used, such that each input unit connects with one trainable weight to the soma and with another trainable weight to the dendrite. Also, for the more complex models (②–⑤ in Fig. 7), we fixed as hyper-parameters the connection strength between the somatic and dendritic compartment as well as the time scale of the filters (see Methods). We found that, with the exception of the models with the parallel motif, all models trained well and achieved above 80% accuracy within 2000 epochs, demonstrating these models are trainable using standard techniques.

Importantly, as with any machine learning model, increasing PRC model complexity increases the number of parameters to be trained and hyperparameters to be tuned. Good machine learning practice involves a systematic methodology for tuning the hyper-parameters (Bergstra et al., 2011, Feurer and Hutter, 2019). Here, however, the particular choice of hyper-parameters was based on our comparison with electrophysiological observations. With these hyper-parameters, good accuracy using the parallel motif was achieved but only in a small portion of the initializations. It is possible that the slow time constant used in the parallel subunits tended to prevent these models from achieving good accuracy. Thus, we have demonstrated that various forms of PRC models can be trained on a memorization task and that combining these models with techniques for hyper-parameter optimization (Bergstra et al., 2011, Feurer and Hutter, 2019) opens the door to comparing model architectures on the basis of training efficacy.

Because few problems solved by machine learning systems or the brain are likely to closely resemble memorization of random patterns of inputs, we next assessed whether our PRC models could be trained to solve a more realistic problem: classification of spoken digits. The Heidelberg digits dataset

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

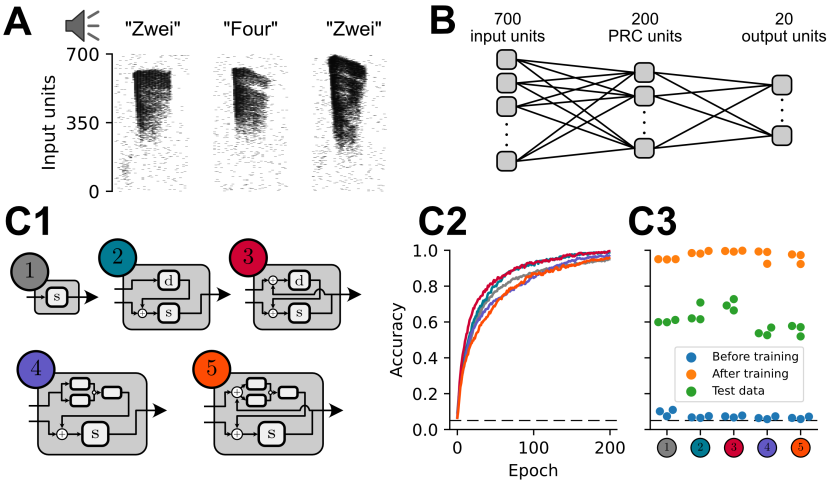

Figure 8: Multi-layered networks of PRC neurons can be trained to recognize spoken digits, and generalize to sound samples from new speakers. **A** Training examples from the Heidelberg spoken digits dataset (Cramer et al., 2020). Audio recordings of digits spoken in English and German were transformed into spiketrains using a detailed model of the inner ear. **B** Multi-layered PRC network architecture, an expanded version of the architecture shown in Fig. 7. **C** Performance of multi-layered PRC networks. Hidden unit architectures (C1) are the same as in Fig. 7E. C2 shows the mean training accuracy across $N = 3$ randomly-initialized networks. Dashed lines indicate chance-level accuracy. C3 shows test set accuracy (green) as well as initial and end training set accuracy for the different models.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contains spiketrains produced by a detailed model of the inner ear in response to audio recordings of the numbers zero to nine spoken by multiple people in two different languages (English and German; Fig. 8A). As a benchmark for assessing the generalization performance of spiking neural network models, the Heidelberg dataset is divided into a standard set of training and test examples, where the test examples include digits spoken by two individuals not included in the training set. To solve this more complex task, we created a set of larger multi-layered PRC network models containing a hidden layer with 200 PRC units in addition to the 700 input units and 20 output units required by the structure of the dataset (Fig. 8B). Training multi-layered networks containing the same set of five PRC models shown in Fig. 7E for 200 epochs consistently improved the training set prediction accuracy from the chance level of 5 % to ≥95 % and as high as 99.5 % for the network containing PRC units which capture the effects of backpropagating action potentials (i.e., ③ in Figs. 7 and 8; analogous to Fig. 6; all values are averages across $N = 3$ randomly-initialized networks; Fig. 7C). The multi-layered PRC networks successfully generalized to new examples, achieving a test set accuracy of 54 % to 70 %, comparable to that previously reported using networks of LIF neurons (Cramer et al., 2020). The prediction accuracy could be further improved by systematic selection of model hyperparameters, including PRC neuron architecture (Bergstra et al., 2011, Feurer and Hutter, 2019). Importantly, these preliminary results provide a clear proof-of-concept that multi-layered networks containing PRC neurons can be trained to solve complex tasks.

Parallel-recurrent cascade models as architectural inductive biases

One of the realities that any machine learning system faces is that no learning agent can be efficient at learning all types of problems and/or tasks. Due to the No Free Lunch Theorem for Optimization (Wolpert and Macready, 1997), we know that there is always a trade-off: for a learning system to achieve good performance in one set of problems, it must sacrifice its performance in other sets of problems. When an engineer introduces some design components into an optimization system that help to bias a learning system towards a particular set of problems, we refer to these components as "inductive biases". Inductive biases are key to developing useful machine learning systems, because without appropriate inductive biases, learning systems will provide mediocre performance on all tasks, as opposed to superior performance on the subset of tasks that we may care about (Hessel et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2019, Goyal and Bengio, 2020). For example, if an engineer creates a machine learning system that is built with an inductive bias to seek out relationships between discrete objects, then that can help the system learn about spatio-visual object relations such as "there is a blue ball to the left of a red cube" (Battaglia et al., 2018, Santoro et al., 2017).

These insights about the importance of inductive biases actually helped to lay the foundations for the modern deep learning approach in AI (Bengio et al., 2007, Goyal and Bengio, 2020). The early proponents of deep learning proposed that the set of problems we most care about in AI are those that humans and/or animals are good at, e.g. image processing, motor control, language comprehension, etc (Bengio et al., 2007). Given this, they argued that machine learning researchers should seek inspiration from real brains in order to find appropriate inductive biases for AI (LeCun et al., 2015). At the time, the principal inductive bias that these researchers were interested in was network depth (hence the name "deep" learning). They believed that the macroscopic architecture of the brain, with multiple stages of processing involved in sensorimotor transformations, provided inductive biases to promote hierarchical representations, which was proposed to solve the type of problems where humans and animals excel (Bengio et al., 2007). In hindsight, it appears that their intuitions were largely correct: deep ANNs have consistently outperformed other types of machine learning approaches in exactly the sort of problem/task domains that humans and animals excel at (LeCun et al., 2015). Moreover, theoretical analyses have provided some explanations for why ANNs with deep architectures are particularly well suited to such applications (Lee et al., 2017, Li and Sompolinsky, 2020).

There is a broader, two-fold point within the story of deep learning and inductive biases. First, it is clear now that the inductive biases of an ANN are determined in large part by architecture, i.e. how the linear and non-linear operations are arranged in a computational graph within the ANN. This is because architecture determines both how information flows through the ANN, but also the shape of the loss landscape, which can directly impact the efficiency with which different types of representations can be learned (Belkin et al., 2019, Du et al., 2019). Second, the success of using the brain's hierarchical structure to inspire the architecture of ANNs demonstrates that, in principle, it can be beneficial for AI to seek inspiration from the brain when seeking new inductive biases (Hassabis et al., 2017). Importantly, PRC models provide a new way of incorporating biological insights into the design of ANN architectures. If we were to replace the standard units of an ANN

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with PRC models based on real neurons, this would represent a major change in the architecture of the ANN, one that may provide useful inductive biases.

Potential advantages of PRC-based inductive biases

When considering the potential inductive biases that PRC models introduce, the natural question is, would these inductive biases actually help or hinder AI? Though it is true that inspiration from brains have provided useful inductive biases for machine learning in the past (LeCun et al., 2015, Hassabis et al., 2017), there is no rule that says that neural phenomena always provide such utility. Indeed, some aspects of physiology may be more related to phylogenetic history and biological constraints than they are to improved learning performance. However, there are a few reasons to think that dendrite inspired PRC models may provide useful inductive biases.

As noted above, research over the last decade has confirmed that network depth is an important architectural consideration for AI (Lee et al., 2017, Li and Sompolinsky, 2020). However, not all depth is equal. Researchers have found that increasing depth is most useful when additional architectural features are included, such as skip connections (Du et al., 2019). If we were to replace the units of an ANN with PRC models based on real neurons that would increase the depth of the network, but in a very particular way. The central question, then, is would the specific form of increased depth that one would obtain from using PRC model units would be helpful?

One reason that PRC models may provide a useful form of depth is that they would help to promote sparsity, which has also been shown to be useful in neural networks (Srivastava et al., 2014). PRC models would help to promote sparsity because they would provide distinct computational subunits with non-linear interactions that could handle different components of a task. For example, it would be possible to have individual dendrites that are responsible for processing distinct types of features, e.g. one set of dendrites for processing facial features, another set for processing body parts, etc. Thus, depending on the input provided, the network could activate only a very sparse set of all the dendrites for processing.

Related to this, there is a growing recognition in the machine learning community that a desirable form of inductive biases for AI would be those that promote the emergence of specialized modules that can be flexibly composed (Goyal et al., 2019, Hinton et al., 2018). This may turn out to be critical to overcoming the limitations of current ANN approaches. Specifically, if an ANN was provided with an architecture that promoted the learn-

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ing of distinct, specialized modules that can be composed, then it should be possible both to learn more intuitive part-whole relationships that capture the underlying structure of objects in the world more accurately (Goyal et al., 2019, Hinton et al., 2018) and to avoid the catastrophic forgetting that can plague standard ANNs (Masse et al., 2018). However, current systems to promote specialized modules are only loosely inspired by real neural circuits. Therefore, an interesting open question is whether PRC models could provide a good mechanism for implementing brain-inspired inductive biases to promote composability. Notably, dendrites have functionally clustered inputs (Kastellakis et al., 2015), synaptic dynamics can perform LNL operations (Rossbroich et al., 2021), and these phenomena interact (Wilson et al., 2016). Given the fact that PRC models would allow an ANN to learn dendrite-like, flexible, non-linear, recurrent interactions between functionally clustered inputs, it is plausible that they could help with composability. Thus, we would argue that future research should investigate the potential for ANNs that use PRC models inspired by sub-cellular dendritic computation to show better specialization and composability, and less catastrophic forgetting.

Discussion

In this article, we have illustrated the capabilities of PRC models to capture the various experimentally observed features of dendritic computation, and discussed how this modeling framework may be key to understand the role of dendrites in learning and neuronal computation. In doing so, we have identified modular, composable connectivity motifs (Fig. 2), with the parallel, recurrent and cascade elements forming the basic building blocks of the framework. We have illustrated how the interaction between sodium spikes in the dendrite and the soma (Golding and Spruston, 1998) can be captured using a recurrent motif of LNL subunits (Fig. 4). Furthermore, we have demonstrated that a parallel motif of LNL subunits within a cascade can reproduce the dependence of NMDA spikes on the activation of sodium and calcium channels (Schiller et al., 2000). Finally, we demonstrated that the interaction between back-propagating action potentials and calcium spikes can be captured by a recurrent motif of LNL subunits with a different set of parameters (Fig. 3) (Larkum et al., 1999, 2004). In closing, we have argued that these PRC models of dendritic computation could have an important role in shaping inductive biases (Section h), and thus contribute to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the optimization of learning capabilities of the brain.

Our LNL models, however, bear some important limitations. Firstly, the set of operations that are possible within the LNL framework correspond to a subset of the operations that are achievable by the type of dynamical systems used in detailed simulations of dendritic integration (Schiller et al., 2000, Hay et al., 2011, Ujfalussy et al., 2018, Beaulieu-Laroche et al., 2018). For instance, modeling the NMDA spike as two LNL units in parallel followed by another nonlinearity ignores the nonlinear impact that sodium ion channels can have on calcium ion channels via rapid increase of the membrane potential. Another element not captured by the phenomenological LNL model is that ion channel time constants almost always depend on the mean depolarization, which implies an adaptive filter instead of a fixed filter assumed in the LNL model. A second important limitation is that we have assumed that the dendrites remain in a fluctuation-driven regime where the net input is low on average but highly variable. If we were to give strong and sustained inputs to the LNL model, these would saturate the nonlinearity and nonlinear transients would disappear. Such a sustained-depolarization regime has been observed in some experiments (Larkum et al., 2004), but it remains to be seen whether these take place *in vivo* or whether homeostatic mechanisms preserve the fluctuation-driven regime (Vogels et al., 2011). One last limitation of our model is that a very high input variability is capable of making nonlinear operations effectively linear (Ujfalussy et al., 2018). In this case, the complex linear-nonlinear structure operates in a way that can actually be captured by an effective model that is entirely linear. Although some *in vivo* manipulation of dendritic inputs (Xu et al., 2012, Gambino et al., 2014, Doron et al., 2020) argue against this point of view, the full relationship between inputs and outputs of neurons in a naturalistic condition is far from being fully known.

We included in this paper some discussion of the potential machine learning applications of PRC models. As we outlined, there are some reasons to believe that dendritic computations may provide useful inductive biases for machine learning systems. We are hopeful that future research will demonstrate this. However, it also has to be recognized that real dendrites may be solving an implementation problem for neurons, i.e. how can you actually integrate thousands of distinct signals in a physical circuit with space and energy constraints? It is possible that this is the problem which dendrites solve for real neurons, and that dendritic computation is, itself, not important at the algorithmic level. Only by exploring the potential advantages of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

training AI systems with PRC models inspired by real neurons will we be able to get some initial insight on this mystery.

In summary, our work shows how PRC models can be used to model subcellular dendritic computation with a computationally tractable approach. This lays the groundwork for future explorations of the algorithmic implications of dendritic computation, both in the brain, and in machine learning applications. We believe that PRC models will help open the door to exploring the true computational power of dendrites.

1. Methods

1.1. Linear-nonlinear subunit

The basic component of the modeling framework presented here is the linear-nonlinear subunit, which receives a net time-varying input $I_i(t)$ and produces an activation $z(t)$ as its output

$$z(t; \kappa_{\text{nlin}}, \kappa_{\text{lin}}, \kappa_{\text{ad}}) = g(a_{\text{nlin}}(t; \kappa_{\text{nlin}}, \kappa_{\text{ad}})) + a_{\text{lin}}(t; \kappa_{\text{lin}}, \kappa_{\text{ad}}) \tag{1}$$

$$a_x(t; \kappa_x, \kappa_{\text{ad}}) = [\kappa_x * (I_{\text{ext}} + I_{\text{ad}})](t) \tag{2}$$

$$I_{\text{ad}}(t; \kappa_{\text{ad}}) = [\kappa_{\text{ad}} * z](t), \tag{3}$$

where $g(\cdot)$ is a nonlinear activation function; $a_{\text{nlin}}(t)$ and $a_{\text{lin}}(t)$ are the pre-activations for the nonlinear and linear parts of the output, respectively; $I_{\text{ext}}(t) = \sum_i I_i(t)$ is the total input from all external sources; $I_{\text{ad}}(t)$ is a recurrent adaptation current; and $(\kappa * x)(t) = \int_{-\infty}^{t} \kappa(t-\tau)x(\tau)d\tau$ denotes the causal convolution of a filter κ with a signal x evaluated at time t. The subunit is parameterized in terms of the activation filters κ_{nlin} and κ_{lin}, the adaptation filter κ_{ad}, and the choice of nonlinear activation function $g(\cdot)$. Except where noted, all filters in this work are defined as exponential functions

$$\kappa(t) = \frac{1}{\tau} e^{\frac{-t}{\tau}} \tag{4}$$

with time constant τ. For models with multiple cellular compartments, $I_{\text{ext}}(t)$ may include inputs from an external source as well as from other compartments (eg, Methods 1.4). Depending on the model, $z(t)$, $a_{\text{nlin}}(t)$, or $a_{\text{lin}}(t)$ may correspond loosely to the voltage of a compartment denoted $V_x(t)$ where x is the name of the compartment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1.2. Two-compartment model subject to a single dendritic input

The two-compartment model with a recurrent connection from the somatic to the dendritic compartment is defined as follows

$$z^{(\mathrm{s})}(t) = g(a^{(\mathrm{s})}_{\mathrm{nlin}}(t;\kappa^{(\mathrm{s})}_{\mathrm{nlin}})) + a^{(\mathrm{s})}_{\mathrm{lin}}(t;\kappa^{(\mathrm{s})}_{\mathrm{lin}}) \tag{5}$$

$$z^{(\mathrm{d})}(t) = \sigma(a^{(\mathrm{d})}_{\mathrm{nlin}}(t;\kappa^{(\mathrm{d})}_{\mathrm{nlin}})) + a^{(\mathrm{d})}_{\mathrm{lin}}(t;\kappa^{(\mathrm{d})}_{\mathrm{lin}}) \tag{6}$$

$$I^{(\mathrm{s})}_{\mathrm{ext}}(t) = I^{(\mathrm{d})}_{\mathrm{ext}}(t) = I_d(t), \tag{7}$$

where the dendritic nonlinearity $\sigma(\cdot)$ is the sigmoid function. $g(\cdot)$ is the somatic spiking nonlinearity, a function which emits a 1 ms square pulse of amplitude $A = 2$ AU when $z^{(\mathrm{s})}(t)$ crosses the somatic spike threshold from below

$$g(t;v) = rect(t) * \sum_{t^{(f)}<t} \delta(t - t^{(f)}) \tag{8}$$

$$rect(t) = \begin{cases} A & \text{for } 0 < t < 1ms \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

where $t^{(f)}$ denotes the time of a threshold crossing. Exponential functions with the following time constants were used for the pre-activation filters: $\tau^{(\mathrm{s})}_{\mathrm{lin}} = 40$ ms, $\tau^{(\mathrm{d})}_{\mathrm{lin}} = 20$ ms, $\tau^{(\mathrm{s})}_{\mathrm{nlin}} = 40$ ms, and $\tau^{(\mathrm{d})}_{\mathrm{nlin}} = 2$ ms. Adaptation filters were set to zero and the terms associated with them were omitted from the above model for simplicity. The activation output $z^{(x)}(t)$ corresponds loosely to the voltage of each compartment; $V_s(t) \equiv z^{(s)}(t)$ and $V_d(t) \equiv z^{(d)}(t)$ are therefore used to refer to these terms in figures and the main text for ease of interpretation.

1.3. Multi-subunit model with parallel processing

The model is composed of three linear-nonlinear subunits which loosely capture the contributions of sodium, calcium, and NMDA voltage-dependent conductances (denoted by the superscripts (1), (2), and (3), respectively) to nonlinear processing of synaptic inputs in a dendritic compartment. Their dynamics are defined as follows

$$z^{(\mathrm{x})}(t) = \sigma\left(a^{(\mathrm{x})}(t;\kappa^{(\mathrm{x})})\right) + a^{(\mathrm{x})}(t;\kappa^{(\mathrm{x})}) \qquad \text{for x} \in \{1,2,3\} \tag{10}$$

$$\kappa^{(\mathrm{x})} = \frac{1}{\tau^{(\mathrm{x})}} e^{\frac{-t}{\tau^{(\mathrm{x})}}}, \tag{11}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with inputs

$$I_{\text{ext}}^{(1)}(t) = I_{\text{ext}}^{(2)}(t) = I_d(t) \tag{12}$$

$$I_{\text{ext}}^{(3)}(t) = z^{(1)}(t) + z^{(2)}(t). \tag{13}$$

In all three subunits, the non-linear and linear pre-activation filters were set to be equal, such that

$$\kappa^{(\text{x})} = \kappa_{\text{lin}}^{(\text{x})} = \kappa_{\text{nlin}}^{(\text{x})} \tag{14}$$

where $\tau^{(1)} = 5\,\text{ms}$, $\tau^{(2)} = 40\,\text{ms}$, and $\tau^{(3)} = 80\,\text{ms}$. The adaptation filters were set to zero and the associated terms dropped for simplicity. The dendritic synaptic-like input current is given by the alpha function

$$I_d(t) = Ate^{\frac{-t}{\tau}} \qquad \text{for } t \geq 0 \tag{15}$$

with amplitude A and time constant $\tau = 2\,\text{ms}$. The voltage output shown in the figures and main text is analogous to the activation output $V_{\text{x}}(t) \equiv z^{(\text{x})}(t)$ for each respective subunit.

1.4. Two-compartment model with bi-directional dendro-somatic interactions

The model with bi-directional dendro-somatic interactions is composed of two reciprocally-connected linear-nonlinear subunits (see Section 1.1) as follows

$$z^{(\text{s})}(t) = \Theta\left(a_{\text{nlin}}^{(\text{s})}(t; \kappa_{\text{nlin}}^{(\text{s})}, \kappa_{\text{ad}}^{(\text{s})})\right) \tag{16}$$

$$I_{\text{ext}}^{(\text{s})}(t) = I_s(t) + z^{(\text{d})}(t) \tag{17}$$

$$z^{(\text{d})}(t) = \sigma\left(a_{\text{nlin}}^{(\text{d})}(t; \kappa_{\text{nlin}}^{(\text{d})})\right) \tag{18}$$

$$I_{\text{ext}}^{(\text{d})}(t) = I_d(t) + z^{(\text{s})}(t), \tag{19}$$

where $\Theta(\cdot)$ is the Heaviside step function and $\sigma(\cdot)$ is the sigmoid function. The nonlinear activation filters $\kappa_{\text{nlin}}^{(x)}$ are defined as exponential functions with $\tau_{\text{nlin}}^{(\text{s})} = 10\,\text{ms}$ and $\tau_{\text{nlin}}^{(\text{d})} = 5\,\text{ms}$. The somatic adaptation filter is defined as

$$\kappa_{\text{ad}}^{(\text{s})}(t) = \begin{cases} 1 & \text{for } t = 0 \\ \frac{1}{\tau_{\text{ad}}} e^{\frac{-t}{\tau_{\text{ad}}}} & \text{for } t > 0 \end{cases} \tag{20}$$

with $\tau_{\text{ad}} = 20\,\text{ms}$. In this model, the linear activation filters $\kappa_{\text{lin}}^{(\text{s})}$ and $\kappa_{(\text{d})}$ are set to zero, along with the dendritic adaptation filter $\kappa_{\text{ad}}^{(\text{d})}$. (The terms

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

associated with these filters have been dropped from the above model definition for simplicity.) $I_s(t)$ and $I_d(t)$ correspond loosely to synaptic inputs to the somatic and dendritic compartments, respectively. The somatic pre-activation $a^{(\mathrm{s})}_{\mathrm{nlin}}$ and the dendritic activation $z^{(\mathrm{d})}$ loosely correspond to the voltage in their respective compartments. For clarity, we use $V_s(t) \equiv a^{(\mathrm{s})}_{\mathrm{nlin}}(t)$ and $V_d(t) \equiv z^{(\mathrm{d})}(t)$ to refer to these quantities in the figures and main text.

1.5. Multi-layered networks of PRC models

Artificial spiking neural networks with a hidden layer of PRC neurons were constructed following the approach of Refs. (Zenke, 2019, Neftci et al., 2019, Cramer et al., 2020). Briefly, spikes in each layer were integrated as exponentially-decaying synaptic currents in each neuron/PRC-compartment in the following layer, which in turn were integrated by the dynamics of the corresponding PRC subunit as $I_{\mathrm{ext}}(t)$. The neurons in the output layer consisted of leaky integrators without threshold or reset. The time-varying voltage of the neurons in the output layer was transformed into a set of class probabilities by applying a softmax operation to the maximum voltage attained by each output unit during each example.

The PRC filters κ used in network models were defined as

$$\kappa_{\mathrm{x}}(t) = e^{\frac{-t}{\tau_{\mathrm{x}}}} \quad \text{for } t \geq 0 \tag{21}$$

where the membrane time constant $\tau_{\mathrm{nlin}} = 10\,\mathrm{ms}$, and the synaptic time constant $\tau_{\mathrm{syn}} = 5\,\mathrm{ms}$. Time constants of sodium, calcium, and NMDA filters took the same values as in Methods 1.3. The linear forward filters κ_{lin} were dropped for simplicity. The adaptation filter in the somatic compartment was defined as

$$\kappa^{(\mathrm{s})}_{\mathrm{ad}} = \begin{cases} 15 & \text{for } t = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

In order to train multi-layered networks of PRC neurons, it was necessary to use a surrogate gradient for the Heaviside step function used to generate spikes in the somatic compartment (since its gradient is zero almost everywhere). Following the approach of (Zenke and Ganguli, 2018, Zenke, 2019), we used the normalized gradient of a sigmoid function to approximate the gradient of the Heaviside function

$$\frac{d}{dx}\Theta(x) \approx \frac{\alpha}{(s|x|+1)^2},$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where α is a proportionality constant and $s = 10$ is a scale parameter that sets the slope of the sigmoid.

All training was carried out using the Adam optimization algorithm with a learning rate of 0.002 and the negative log likelihood of making a correct class prediction as the loss function.

1.6. Numerical methods

Simulations were implemented in Matlab and Python 3.8 using NumPy 1.18.5, SciPy 1.5.0, and ez-ephys 0.4.2. Figures were prepared in Python using Matplotlib 3.2.2, Jupyter 1.0.0, and ez-ephys. Code is available at .

References

Ariav G, Polsky A, Schiller J (2003) Submillisecond precision of the input-output transformation function mediated by fast sodium dendritic spikes in basal dendrites of ca1 pyramidal neurons. Journal of Neuroscience 23:7750–7758.

Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R et al. (2018) Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261 .

Beaulieu-Laroche L, Toloza EH, Van der Goes MS, Lafourcade M, Barnagian D, Williams ZM, Eskandar EN, Frosch MP, Cash SS, Harnett MT (2018) Enhanced dendritic compartmentalization in human cortical neurons. Cell 175:643–651.

Belkin M, Hsu D, Ma S, Mandal S (2019) Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences 116:15849–15854.

Bengio Y, LeCun Y et al. (2007) Scaling learning algorithms towards ai. Large-scale kernel machines 34:1–41.

Bergstra J, Bardenet R, Bengio Y, Kégl B (2011) Algorithms for hyper-parameter optimization. Advances in neural information processing systems 24.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bird AD, Cuntz H (2020) Dendritic normalisation improves learning in sparsely connected artificial neural networks. bioRxiv .

Brunel N, Hakim V, Richardson MJ (2014) Single neuron dynamics and computation. Current opinion in neurobiology 25:149–155.

Cook EP, Johnston D (1997) Active dendrites reduce location-dependent variability of synaptic input trains. Journal of neurophysiology 78:2116–2128.

Cramer B, Stradmann Y, Schemmel J, Zenke F (2020) The heidelberg spiking data sets for the systematic evaluation of spiking neural networks. IEEE Transactions on Neural Networks and Learning Systems .

Deitcher Y, Eyal G, Kanari L, Verhoog MB, Atenekeng Kahou GA, Mansvelder HD, De Kock CP, Segev I (2017) Comprehensive morpho-electrotonic analysis shows 2 distinct classes of l2 and l3 pyramidal neurons in human temporal cortex. Cerebral Cortex 27:5398–5414.

Doron G, Shin JN, Takahashi N, Drüke M, Bocklisch C, Skenderi S, de Mont L, Toumazou M, Ledderose J, Brecht M et al. (2020) Perirhinal input to neocortical layer 1 controls learning. Science 370.

Du S, Lee J, Li H, Wang L, Zhai X (2019) Gradient descent finds global minima of deep neural networks In International Conference on Machine Learning, pp. 1675–1685. PMLR.

Eyal G, Mansvelder HD, de Kock CP, Segev I (2014) Dendrites impact the encoding capabilities of the axon. Journal of Neuroscience 34:8063–8071.

Feurer M, Hutter F (2019) Hyperparameter optimization In Automated machine learning, pp. 3–33. Springer, Cham.

Francioni V, Harnett MT (2021) Rethinking single neuron electrical compartmentalization: dendritic contributions to network computation in vivo. Neuroscience .

Gambino F, Pagès S, Kehayas V, Baptista D, Tatti R, Carleton A, Holtmaat A (2014) Sensory-evoked ltp driven by dendritic plateau potentials in vivo. Nature 515:116–119.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gasparini S, Migliore M, Magee JC (2004) On the initiation and propagation of dendritic spikes in ca1 pyramidal neurons. Journal of Neuroscience 24:11046–11056.

Gerstner W, Kistler WM, Naud R, Paninski L (2014) Neuronal dynamics: From single neurons to networks and models of cognition Cambridge University Press.

Gidon A, Zolnik TA, Fidzinski P, Bolduan F, Papoutsi A, Poirazi P, Holtkamp M, Vida I, Larkum ME (2020) Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science 367:83–87.

Golding NL, Spruston N (1998) Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal ca1 pyramidal neurons. Neuron 21:1189–1200.

Goyal A, Bengio Y (2020) Inductive biases for deep learning of higher-level cognition. arXiv preprint arXiv:2011.15091 .

Goyal A, Lamb A, Hoffmann J, Sodhani S, Levine S, Bengio Y, Schölkopf B (2019) Recurrent independent mechanisms. arXiv preprint arXiv:1909.10893 .

Hassabis D, Kumaran D, Summerfield C, Botvinick M (2017) Neuroscience-inspired artificial intelligence. Neuron 95:245–258.

Hay E, Hill S, Schürmann F, Markram H, Segev I (2011) Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. PLoS Comput Biol 7:e1002107.

Herz AV, Gollisch T, Machens CK, Jaeger D (2006) Modeling single-neuron dynamics and computations: a balance of detail and abstraction. science 314:80–85.

Hessel M, van Hasselt H, Modayil J, Silver D (2019) On inductive biases in deep reinforcement learning. arXiv preprint arXiv:1907.02908 .

Hinton GE, Sabour S, Frosst N (2018) Matrix capsules with em routing In International conference on learning representations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of physiology 117:500–544.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology 195:215–243.

Jolivet R, Kobayashi R, Rauch A, Naud R, Shinomoto S, Gerstner W (2008) A benchmark test for a quantitative assessment of simple neuron models. Journal of neuroscience methods 169:417–424.

Jolivet R, Rauch A, Lüscher HR, Gerstner W (2006) Predicting spike timing of neocortical pyramidal neurons by simple threshold models. Journal of computational neuroscience 21:35–49.

Jones IS, Kording KP (2021) Do biological constraints impair dendritic computation? arXiv e-prints pp. arXiv–2103.

Kalmbach BE, Gray R, Johnston D, Cook EP (2017) Systems-based analysis of dendritic nonlinearities reveals temporal feature extraction in mouse l5 cortical neurons. Journal of neurophysiology 117:2188–2208.

Kastellakis G, Cai DJ, Mednick SC, Silva AJ, Poirazi P (2015) Synaptic clustering within dendrites: an emerging theory of memory formation. Progress in neurobiology 126:19–35.

Kobayashi R, Tsubo Y, Shinomoto S (2009) Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. Frontiers in computational neuroscience 3:9.

Koch C (2004) Biophysics of computation: information processing in single neurons Oxford university press.

Krichmar JL, Nasuto SJ, Scorcioni R, Washington SD, Ascoli GA (2002) Effects of dendritic morphology on ca3 pyramidal cell electrophysiology: a simulation study. Brain research 941:11–28.

Kuffler SW (1953) Discharge patterns and functional organization of mammalian retina. Journal of neurophysiology 16:37–68.

Larkum ME, Senn W, Lüscher HR (2004) Top-down dendritic input increases the gain of layer 5 pyramidal neurons. Cerebral cortex 14:1059–1070.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Larkum ME, Zhu JJ (2002) Signaling of layer 1 and whisker-evoked ca2+ and na+ action potentials in distal and terminal dendrites of rat neocortical pyramidal neurons in vitro and in vivo. Journal of neuroscience 22:6991–7005.

Larkum ME, Zhu JJ, Sakmann B (1999) A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature 398:338–341.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. nature 521:436–444.

Lee J, Bahri Y, Novak R, Schoenholz SS, Pennington J, Sohl-Dickstein J (2017) Deep neural networks as gaussian processes. arXiv preprint arXiv:1711.00165 .

Legenstein R, Maass W (2011) Branch-specific plasticity enables self-organization of nonlinear computation in single neurons. Journal of Neuroscience 31:10787–10802.

Li Q, Sompolinsky H (2020) Statistical mechanics of deep linear neural networks: The back-propagating renormalization group. arXiv preprint arXiv:2012.04030 .

Magee JC, Carruth M (1999) Dendritic voltage-gated ion channels regulate the action potential firing mode of hippocampal ca1 pyramidal neurons. Journal of neurophysiology 82:1895–1901.

Mainen ZF, Sejnowski TJ (1996) Influence of dendritic structure on firing pattern in model neocortical neurons. Nature 382:363–366.

Masse NY, Grant GD, Freedman DJ (2018) Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. Proceedings of the National Academy of Sciences 115:E10467–E10475.

McFarland JM, Cui Y, Butts DA (2013) Inferring nonlinear neuronal computation based on physiologically plausible inputs. PLoS Comput Biol 9:e1003143.

McKenna TM, Davis JL, Zornetzer SF (2014) Single neuron computation Academic Press.

Mel BW (1993) Synaptic integration in an excitable dendritic tree. Journal of neurophysiology 70:1086–1101.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mensi S, Naud R, Pozzorini C, Avermann M, Petersen CC, Gerstner W (2012) Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. Journal of neurophysiology 107:1756–1775.

Naud R, Bathellier B, Gerstner W (2014) Spike-timing prediction in cortical neurons with active dendrites. Frontiers in computational neuroscience 8:90.

Naud R, Payeur A, Longtin A (2017) Noise gated by dendrosomatic interactions increases information transmission. Phys. Rev. X 7:031045.

Neftci EO, Mostafa H, Zenke F (2019) Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. IEEE Signal Processing Magazine 36:51–63.

Papoutsi A, Kastellakis G, Poirazi P (2017) Basal tree complexity shapes functional pathways in the prefrontal cortex. Journal of neurophysiology 118:1970–1983.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky E (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. Journal of Neuroscience 25:11003–11013.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky E, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454:995–999.

Pinsky PF, Rinzel J (1994) Intrinsic and network rhythmogenesis in a reduced traub model for ca3 neurons. Journal of computational neuroscience 1:39–60.

Poirazi P, Brannon T, Mel BW (2003) Pyramidal neuron as two-layer neural network. Neuron 37:989–999.

Poirazi P, Mel BW (2001) Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron 29:779–796.

Poirazi P, Papoutsi A (2020) Illuminating dendritic function with computational models. Nature Reviews Neuroscience 21:303–321.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pozzorini C, Naud R, Mensi S, Gerstner W (2013) Temporal whitening by power-law adaptation in neocortical neurons. Nature neuroscience 16:942–948.

Psarrou M, Stefanou SS, Papoutsi A, Tzilivaki A, Cutsuridis V, Poirazi P (2014) A simulation study on the effects of dendritic morphology on layer v prefrontal pyramidal cell firing behavior. Frontiers in cellular neuroscience 8:287.

Rall W (1960) Membrane potential transients and membrane time constant of motoneurons. Experimental neurology 2:503–532.

Richards BA, Lillicrap TP, Beaudoin P, Bengio Y, Bogacz R, Christensen A, Clopath C, Costa RP, de Berker A, Ganguli S et al. (2019) A deep learning framework for neuroscience. Nature neuroscience 22:1761–1770.

Rossbroich J, Trotter D, Beninger J, Tóth K, Naud R (2021) Linear-nonlinear cascades capture synaptic dynamics. PLOS Computational Biology 17:1–27.

Santoro A, Raposo D, Barrett DG, Malinowski M, Pascanu R, Battaglia P, Lillicrap T (2017) A simple neural network module for relational reasoning. arXiv preprint arXiv:1706.01427 .

Schaefer AT, Larkum ME, Sakmann B, Roth A (2003) Coincidence detection in pyramidal neurons is tuned by their dendritic branching pattern. Journal of neurophysiology 89:3143–3154.

Schiller J, Major G, Koester HJ, Schiller Y (2000) Nmda spikes in basal dendrites of cortical pyramidal neurons. Nature 404:285–289.

Shai AS, Anastassiou CA, Larkum ME, Koch C (2015) Physiology of layer 5 pyramidal neurons in mouse primary visual cortex: coincidence detection through bursting. PLoS Comput Biol 11:e1004090.

Sinz FH, Pitkow X, Reimer J, Bethge M, Tolias AS (2019) Engineering a less artificial intelligence. Neuron 103:967–979.

Smith SL, Smith IT, Branco T, Häusser M (2013) Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. Nature 503:115–120.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research 15:1929–1958.

Stein RB (1965) A theoretical analysis of neuronal variability. Biophysical Journal 5:173–194.

Taylor WR, He S, Levick WR, Vaney DI (2000) Dendritic computation of direction selectivity by retinal ganglion cells. Science 289:2347–2350.

Teeter C, Iyer R, Menon V, Gouwens N, Feng D, Berg J, Szafer A, Cain N, Zeng H, Hawrylycz M et al. (2018) Generalized leaky integrate-and-fire models classify multiple neuron types. Nature communications 9:1–15.

Truccolo W, Donoghue JA, Hochberg LR, Eskandar EN, Madsen JR, Anderson WS, Brown EN, Halgren E, Cash SS (2011) Single-neuron dynamics in human focal epilepsy. Nature neuroscience 14:635.

Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of neurophysiology 93:1074–1089.

Tzilivaki A, Kastellakis G, Poirazi P (2019) Challenging the point neuron dogma: Fs basket cells as 2-stage nonlinear integrators. Nature communications 10:1–14.

Ujfalussy BB, Makara JK, Lengyel M, Branco T (2018) Global and multiplexed dendritic computations under in vivo-like conditions. Neuron 100:579–592.

Urbanczik R, Senn W (2014) Learning by the dendritic prediction of somatic spiking. Neuron 81:521–528.

Vetter P, Roth A, Hausser M (2001) Propagation of action potentials in dendrites depends on dendritic morphology. Journal of neurophysiology 85:926–937.

Vintch B, Movshon JA, Simoncelli EP (2015) A convolutional subunit model for neuronal responses in macaque v1. Journal of Neuroscience 35:14829–14841.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437091; this version posted July 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W (2011) Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science 334:1569–1573.

Weber AI, Pillow JW (2017) Capturing the dynamical repertoire of single neurons with generalized linear models. Neural computation 29:3260–3289.

Wilson DE, Whitney DE, Scholl B, Fitzpatrick D (2016) Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. Nature neuroscience 19:1003–1009.

Wolpert D, Macready W (1997) No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1:67–82.

Xu Nl, Harnett MT, Williams SR, Huber D, O'Connor DH, Svoboda K, Magee JC (2012) Nonlinear dendritic integration of sensory and motor input during an active sensing task. Nature 492:247–251.

Zenke F (2019) Spytorch.

Zenke F, Ganguli S (2018) Superspike: Supervised learning in multilayer spiking neural networks. Neural computation 30:1514–1541.