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

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

[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] *Mila, Montréal, QC, Canada*
[d] *Montreal Neurological Institute, Montréal, QC, Canada*
[e] *Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada*
[f] *School of Computer Science, McGill University, Montréal, QC, Canada*
[g] *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 show that it is possible to capture these additional phenomena using *parallel, recurrent* cascade models, wherein an individual neuron is modelled as a cascade of parallel linear and non-linear operations that can be connected

[*] 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)

recurrently, akin to a multi-layer, recurrent, artificial neural network. 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

---

1. 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 [1, 2, 3, 4, 5, 6]. 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 [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. 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 and the requirement for computationally demanding numerical simulations, 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?" [7, 21]. 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 [21, 22].

To this end, previous work has explored the use of "cascade models" to capture dendritic computation [23, 24, 25, 26]. Cascade models use a cascade of linear and non-linear operations, which are effectively tree-structured, feedforward, multi-layer artificial neural networks (ANNs) [7]. 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) [23, 24, 26]. 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 and easy to fit to data. Moreover, because these models are fully differentiable, they could easily be incorporated into machine learning approaches that utilize gradient descent for model optimization [21]. 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) [21, 22]. Thus, we conclude this paper by providing some speculation as to the possible advantages for AI of dendritic PRC models. Altogether, this work 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.

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

Models made of LNL operations have had a long history in systems neuroscience, where such models were conceptually implied by early work on retinal ganglion cells [27] and cortical cells [28]. 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

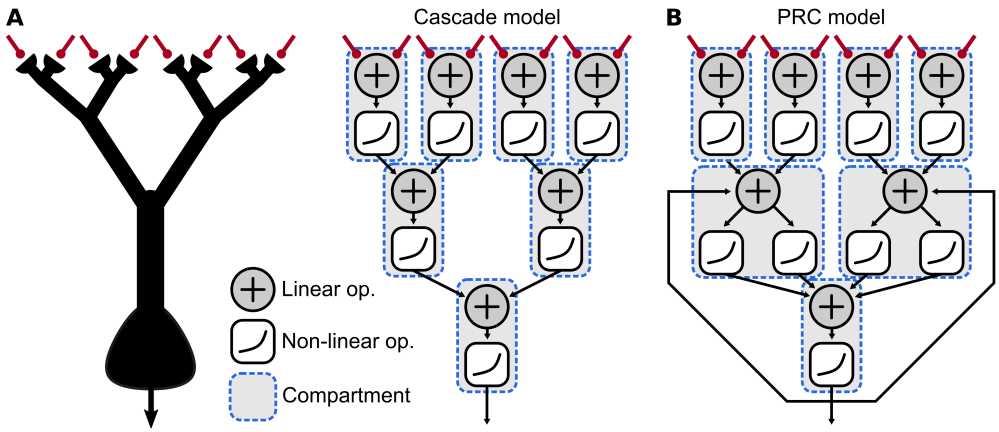

Figure 1: Illustration of cascade models and PRC models: **(A)** Dendritic computation can be modelled using a cascade of linear and non-linear operations [23, 24, 26]. **(B)** Cascade models can be extended with the use of parallel pathways and recurrence in the operations, which can enable the modeling of more complicated regenerative phenomena.

the accuracy of these models were the inclusion of spike-triggered adaptation [29, 30], the composition of linear-nonlinear (LNL) operations in a cascade [31, 32], and the addition of recurrent interactions between linear-nonlinear subunits [33, 34]. 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 the output of the nonlinearity to capture spike-triggered adaptation (Fig. 2A; Methods 5.1). 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 [35]. 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.

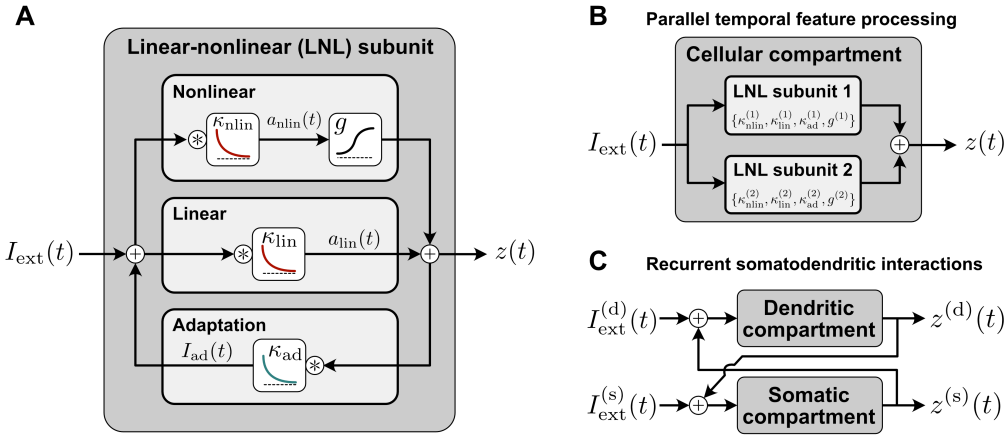

Figure 2: Linear-nonlinear operations and interconnection motifs. **A** The linear-nonlinear (LNL) subunit. This architecture combines linear filtering of an input $I_{\mathrm{ext}}(t)$ (middle) with a nonlinear readout (top, g) and feedback (bottom, κ_{ad}), which together flexibly capture the passive 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.

2.1. Somatic spikes

Since the pioneering work of Richard Stein [36], 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 [37]. Adding multiple time-scales to the kernel for spike-triggered adaptation makes these models highly predictive of the responses a variety of cell types [38, 39, 40, 41, 42]. Furthermore, considering smooth nonlinearities and surrogate gradients can ensure that the LNL unit remains differentiable.

2.2. 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 [43, 44, 23], then using models capturing the dynamics of dendrites and their intrinsic nonlinear properties [45, 25, 26]. These contributions are examples of what we call PRC models because they involve cascade of nonlinearities, but they leave out the recurrent motifs. Also, most have not considered the parallel processing introduced by Ujfalussy et al. (2018) [26]. Recurrence was, however, part of other efforts focusing on simplified models of the interaction between somata and dendrites [46, 47, 48, 49]. Thus our goal in this section is to unify these complementary perspectives.

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 [50, 51, 52, 53]. 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

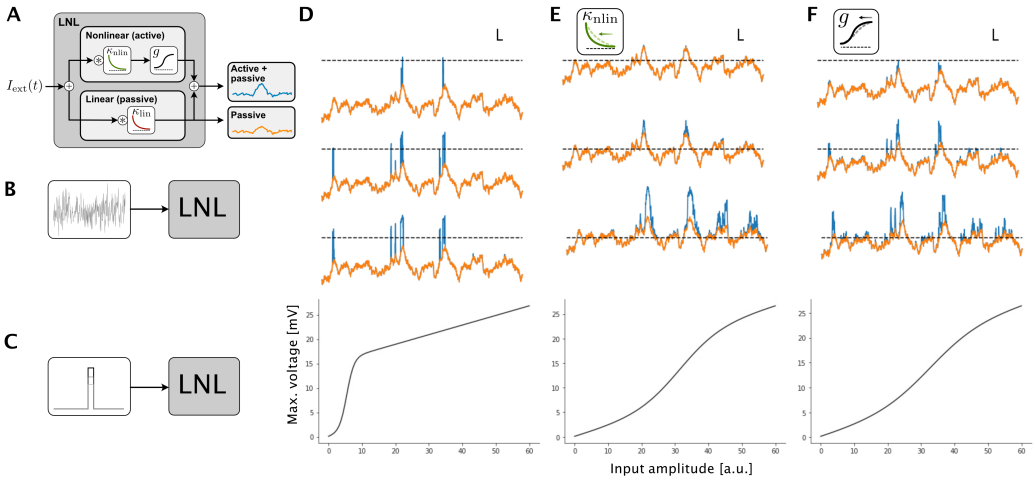

Figure 3: Effects of changing parameters in the linear-nonlinear model. **A** A linear nonlinear model is simulated to produce a nonlinear component (blue) which is added to a passive component (orange). As input to the model, we consider **B** a noisy time-dependent signal representing a bombardment of asynchronous synaptic currents, and **C** a short pulse. **D** Response to noisy inputs having three different baselines (top three traces, lower baseline is topmost trace). Maximum amplitude of response as a function of pulse amplitude (bottom). **E** Effects of increasing the timescale of the linear filter. **F** Effects of decreasing the sensitivity of the nonlinearity.

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 [25].

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 [54, 51, 55, 56] or, with an even longer timescale, NMDA spikes [53]. 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 linear filter, we only changed the gain of the nonlinearity, then the spikes had a more variable amplitude and duration. In these simulations, we have not included an adaptation-like recurrence, which can control the degree of variability in amplitude and duration of the dendritic spikes. Thus, altogether, by changing the parameters of the LNL model, we can simulate some basic electrophysiological features of various types of dendritic spikes.

2.3. 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) [50] 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 *before* triggering an action potential in the

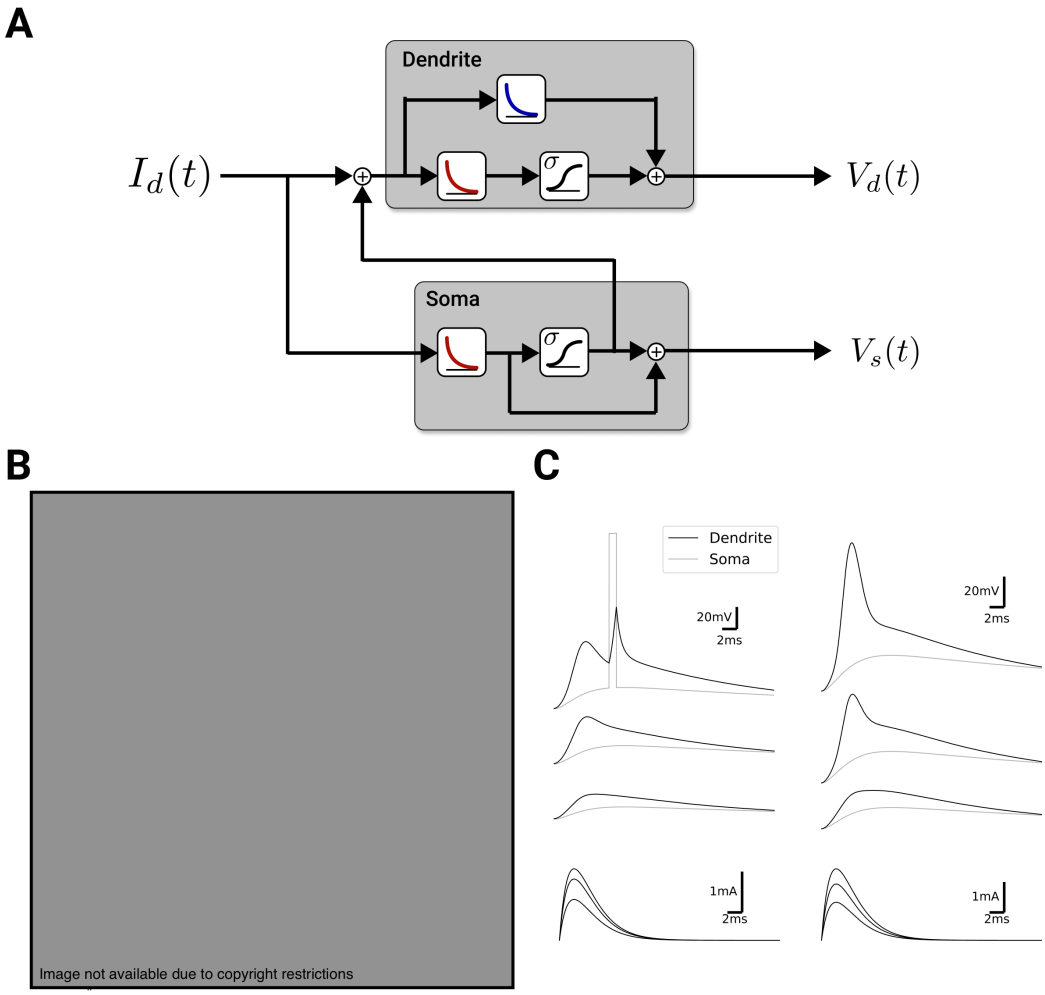

Figure 4: A recurrent motif of linear-nonlinear models for the dendritic sodium spikes. **A** Schematic of the model: A dendritic current (I_d) impinges on two LNL subunits with recurrent interactions, one corresponding to a dendrite, another corresponding to the soma. When the somatic compartment reaches 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 highest input). Two exemplars are shown in different columns. Reproduced from Golding et al. (1998) Fig. 1 [50]. **C** Model responses showing three amplitude levels of a synaptic-like inputs in the model shown in A. **D** 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.

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.

2.4. Dendritic NMDA spikes

Next we considered the electrophysiological recordings of NMDA spikes reported in Schiller et al. (2000) [53]. 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. 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).

2.5. 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 [54, 55]. 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

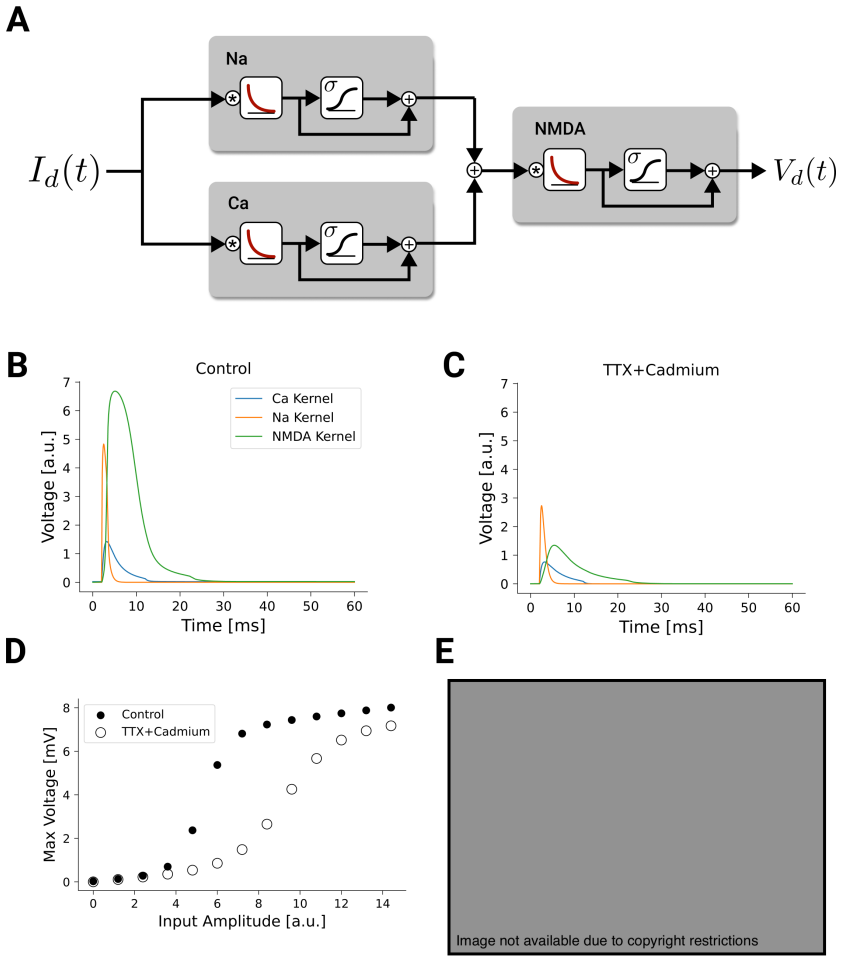

Figure 5: Linear nonlinear model of a NMDA spike as a combination of cascade and parallel processing. **A** Schematic of the model: current impinging on the dendrite is passed through two LNL operations in parallel before feeding into another LNL operation to produce an NMDA spike. **B** Response of 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 E (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 [53].

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 5.4) 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 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 [54]. 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 [54]. 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 [57]. Together, these simulations add to previous efforts [48] in showing that the PRC framework is able to capture features of dendritic excitability and somatodendritic interactions.

3. 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 [58] 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 [26, 59]. 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 [60, 59].

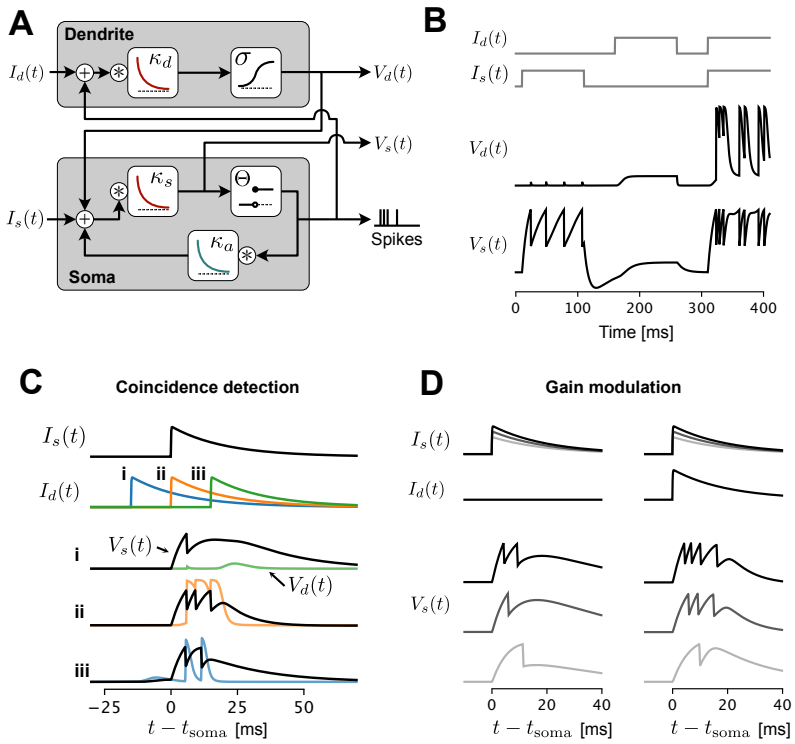

Figure 6: A recurrent cascade model of the interaction between the back-propagating action potential and calcium spikes. **A** Schematic of the model: external current impinging on the dendrite (I_d) and somae (I_s) each engage a LNL operation. The output of the nonlinearities is combined to the input 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 simulataneously 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 [54]. **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 [57].

3.1. 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 [61], 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 [62, 63]. 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" [64, 65].

These insights about the importance of inductive biases actually helped to lay the foundations for the modern deep learning approach in AI [66, 63]. 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 [66]. Given this, they argued that machine learning researchers should seek inspiration from real brains in order to find appropriate inductive biases for AI [67]. 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 an inductive biases to promote hierarchical representations, which they proposed were well suited to solving the sorts of problems humans and animals are good at [66]. 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 [67]. Moreover, theoretical analyses have provided some explanations for why ANNs with deep architectures are particularly well suited to such applications [68, 69].

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 [70, 71]. 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 [72]. 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 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.

3.2. 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 [67, 72], 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 [68, 69]. 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 [71]. 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 a useful in neural networks [73]. PRC models would help to promote sparsity because they would provide distinct computational sub-units 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 [74, 75]. 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 learning 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 [74, 75] and to avoid the catastrophic forgetting that can plague standard ANNs [76]. 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 [77], synaptic dynamics can perform LNL operations [78], and these phenomena interact [79]. 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.

4. Conclusion

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 [50] 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 [53]. 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) [54, 57]. In closing, we have argued that these PRC models of dendritic computation could have an important role in shaping inductive biases (Section 3), and thus contribute to 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 [53, 80, 26, 81]. 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 [57], but it remains to be seen whether these take place *in vivo* or whether homeostatic mechanisms preserve the fluctuation-driven regime [82]. One last limitation of our model is that a very high input variability is capable of making nonlinear operations effectively linear [26]. 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 [56, 83, 84] 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 demon-

strate 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 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.

5. Methods

5.1. 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 5.4). 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.

5.2. 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

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

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

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

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 $z^{(\text{s})}(t)$ 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: $\tau^{(\text{s})}_{\text{lin}} = 40$ ms, $\tau^{(\text{d})}_{\text{lin}} = 20$ ms, $\tau^{(\text{s})}_{\text{nlin}} = 40$ ms, and $\tau^{(\text{d})}_{\text{nlin}} = 2$ 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.

5.3. 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

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

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

$$I^{(1)}_{\mathrm{ext}}(t) = I^{(2)}_{\mathrm{ext}}(t) = I_d(t) \tag{12}$$

$$I^{(3)}_{\mathrm{ext}}(t) = z^{(1)}(t) + z^{(2)}(t). \tag{13}$$

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

$$\kappa^{(\mathrm{x})} = \kappa^{(\mathrm{x})}_{\mathrm{lin}} = \kappa^{(\mathrm{x})}_{\mathrm{nlin}} \tag{14}$$

where $\tau^{(1)} = 5\,\mathrm{ms}$, $\tau^{(2)} = 40\,\mathrm{ms}$, and $\tau^{(3)} = 80\,\mathrm{ms}$. 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 $\tau = 2\,\mathrm{ms}$. The voltage output shown in the figures and main text is analogous to the activation output $V_{\mathrm{x}}(t) \equiv z^{(x)}(t)$ for each respective subunit.

5.4. 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 5.1) as follows

$$z^{(\mathrm{s})}(t) = \Theta(a^{(\mathrm{s})}_{\mathrm{nlin}}(t;\kappa^{(\mathrm{s})}_{\mathrm{nlin}},\kappa^{(\mathrm{s})}_{\mathrm{ad}})) \tag{16}$$

$$I^{(\mathrm{s})}_{\mathrm{ext}}(t) = I_s(t) + z^{(\mathrm{d})}(t) \tag{17}$$

$$z^{(\mathrm{d})}(t) = \sigma(a^{(\mathrm{d})}_{\mathrm{nlin}}(t;\kappa^{(\mathrm{d})}_{\mathrm{nlin}})) \tag{18}$$

$$I^{(\mathrm{d})}_{\mathrm{ext}}(t) = I_d(t) + z^{(\mathrm{s})}(t), \tag{19}$$

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

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

with $\tau_{\mathrm{ad}} = 20\,\mathrm{ms}$. In this model, the linear activation filters $\kappa_{\mathrm{lin}}^{(\mathrm{s})}$ and $\kappa_{(\mathrm{d})}$ are set to zero, along with the dendritic adaptation filter $\kappa_{\mathrm{ad}}^{(\mathrm{d})}$. (The terms 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 $a_{\mathrm{nlin}}^{(\mathrm{s})}$ and the dendritic activation $z^{(\mathrm{d})}$ loosely correspond to the voltage in their respective compartments. For clarity, we use $V_s(t) \equiv a_{\mathrm{nlin}}^{(\mathrm{s})}(t)$ and $V_d(t) \equiv z^{(\mathrm{d})}(t)$ to refer to these quantities in the figures and main text.

5.5. 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 https://github.com/nauralcodinglab/linear-nonlinear-dendrites.

References

[1] N. Brunel, V. Hakim, M. J. Richardson, Single neuron dynamics and computation, Current opinion in neurobiology 25 (2014) 149–155.

[2] A. V. Herz, T. Gollisch, C. K. Machens, D. Jaeger, Modeling single-neuron dynamics and computations: a balance of detail and abstraction, science 314 (2006) 80–85.

[3] A. L. Hodgkin, A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, The Journal of physiology 117 (1952) 500–544.

[4] C. Koch, Biophysics of computation: information processing in single neurons, Oxford university press, 2004.

[5] T. M. McKenna, J. L. Davis, S. F. Zornetzer, Single neuron computation, Academic Press, 2014.

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

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

[8] Z. F. Mainen, T. J. Sejnowski, Influence of dendritic structure on firing pattern in model neocortical neurons, Nature 382 (1996) 363–366.

[9] A. T. Schaefer, M. E. Larkum, B. Sakmann, A. Roth, Coincidence detection in pyramidal neurons is tuned by their dendritic branching pattern, Journal of neurophysiology 89 (2003) 3143–3154.

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

[11] A. S. Shai, C. A. Anastassiou, M. E. Larkum, C. Koch, Physiology of layer 5 pyramidal neurons in mouse primary visual cortex: coincidence detection through bursting, PLoS Comput Biol 11 (2015) e1004090.

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

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

[14] J. L. Krichmar, S. J. Nasuto, R. Scorcioni, S. D. Washington, G. A. Ascoli, Effects of dendritic morphology on ca3 pyramidal cell electrophysiology: a simulation study, Brain research 941 (2002) 11–28.

[15] E. P. Cook, D. Johnston, Active dendrites reduce location-dependent variability of synaptic input trains, Journal of neurophysiology 78 (1997) 2116–2128.

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

[17] G. Ariav, A. Polsky, J. Schiller, 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 (2003) 7750–7758.

[18] G. Eyal, H. D. Mansvelder, C. P. de Kock, I. Segev, Dendrites impact the encoding capabilities of the axon, Journal of Neuroscience 34 (2014) 8063–8071.

[19] Y. Deitcher, G. Eyal, L. Kanari, M. B. Verhoog, G. A. Atenekeng Kahou, H. D. Mansvelder, C. P. De Kock, I. Segev, Comprehensive morpho-electrotonic analysis shows 2 distinct classes of l2 and l3 pyramidal neurons in human temporal cortex, Cerebral Cortex 27 (2017) 5398–5414.

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

[21] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. de Berker, S. Ganguli, et al., A deep learning framework for neuroscience, Nature neuroscience 22 (2019) 1761–1770.

[22] F. H. Sinz, X. Pitkow, J. Reimer, M. Bethge, A. S. Tolias, Engineering a less artificial intelligence, Neuron 103 (2019) 967–979.

[23] P. Poirazi, T. Brannon, B. W. Mel, Pyramidal neuron as two-layer neural network, Neuron 37 (2003) 989–999.

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

[25] B. E. Kalmbach, R. Gray, D. Johnston, E. P. Cook, Systems-based analysis of dendritic nonlinearities reveals temporal feature extraction in mouse l5 cortical neurons, Journal of neurophysiology 117 (2017) 2188–2208.

[26] B. B. Ujfalussy, J. K. Makara, M. Lengyel, T. Branco, Global and multiplexed dendritic computations under in vivo-like conditions, Neuron 100 (2018) 579–592.

[27] S. W. Kuffler, Discharge patterns and functional organization of mammalian retina, Journal of neurophysiology 16 (1953) 37–68.

[28] D. H. Hubel, T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, The Journal of physiology 195 (1968) 215–243.

[29] J. W. Pillow, L. Paninski, V. J. Uzzell, E. P. Simoncelli, E. Chichilnisky, Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model, Journal of Neuroscience 25 (2005) 11003–11013.

[30] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, E. N. Brown, A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects, Journal of neurophysiology 93 (2005) 1074–1089.

[31] B. Vintch, J. A. Movshon, E. P. Simoncelli, A convolutional subunit model for neuronal responses in macaque v1, Journal of Neuroscience 35 (2015) 14829–14841.

[32] J. M. McFarland, Y. Cui, D. A. Butts, Inferring nonlinear neuronal computation based on physiologically plausible inputs, PLoS Comput Biol 9 (2013) e1003143.

[33] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, E. P. Simoncelli, Spatio-temporal correlations and visual signalling in a complete neuronal population, Nature 454 (2008) 995–999.

[34] W. Truccolo, J. A. Donoghue, L. R. Hochberg, E. N. Eskandar, J. R. Madsen, W. S. Anderson, E. N. Brown, E. Halgren, S. S. Cash, Single-neuron dynamics in human focal epilepsy, Nature neuroscience 14 (2011) 635.

[35] W. R. Taylor, S. He, W. R. Levick, D. I. Vaney, Dendritic computation of direction selectivity by retinal ganglion cells, Science 289 (2000) 2347–2350.

[36] R. B. Stein, A theoretical analysis of neuronal variability, Biophysical Journal 5 (1965) 173–194.

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

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

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

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

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

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

[43] B. W. Mel, Synaptic integration in an excitable dendritic tree, Journal of neurophysiology 70 (1993) 1086–1101.

[44] P. Poirazi, B. W. Mel, Impact of active dendrites and structural plasticity on the memory capacity of neural tissue, Neuron 29 (2001) 779–796.

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

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

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

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

[49] R. Naud, A. Payeur, A. Longtin, Noise gated by dendrosomatic interactions increases information transmission, Phys. Rev. X 7 (2017) 031045. URL: https://link.aps.org/doi/10.1103/PhysRevX.7.031045. doi:10.1103/PhysRevX.7.031045.

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

[51] M. E. Larkum, J. J. Zhu, 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 (2002) 6991–7005.

[52] S. L. Smith, I. T. Smith, T. Branco, M. Häusser, Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo, Nature 503 (2013) 115–120.

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

[54] M. E. Larkum, J. J. Zhu, B. Sakmann, A new cellular mechanism for coupling inputs arriving at different cortical layers, Nature 398 (1999) 338–341.

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

[56] N.-l. Xu, M. T. Harnett, S. R. Williams, D. Huber, D. H. O'Connor, K. Svoboda, J. C. Magee, Nonlinear dendritic integration of sensory and motor input during an active sensing task, Nature 492 (2012) 247–251.

[57] M. E. Larkum, W. Senn, H.-R. Lüscher, Top-down dendritic input increases the gain of layer 5 pyramidal neurons, Cerebral cortex 14 (2004) 1059–1070.

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

[59] I. S. Jones, K. P. Kording, Do biological constraints impair dendritic computation?, arXiv e-prints (2021) arXiv–2103.

[60] A. D. Bird, H. Cuntz, Dendritic normalisation improves learning in sparsely connected artificial neural networks, bioRxiv (2020).

[61] D. Wolpert, W. Macready, No free lunch theorems for optimization 1 (????) 67–82. URL: http://ieeexplore.ieee.org/document/585893/. doi:10.1109/4235.585893.

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

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

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

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

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

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

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

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

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

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

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

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

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

[75] G. E. Hinton, S. Sabour, N. Frosst, Matrix capsules with em routing, in: International conference on learning representations, 2018.

[76] N. Y. Masse, G. D. Grant, D. J. Freedman, Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization, Proceedings of the National Academy of Sciences 115 (2018) E10467–E10475.

[77] G. Kastellakis, D. J. Cai, S. C. Mednick, A. J. Silva, P. Poirazi, Synaptic clustering within dendrites: an emerging theory of memory formation, Progress in neurobiology 126 (2015) 19–35.

[78] J. Rossbroich, D. Trotter, J. Beninger, K. Tóth, R. Naud, Linear-nonlinear cascades capture synaptic dynamics, PLOS Computational Biology 17 (2021) 1–27. URL: `https://doi.org/10.1371/journal.pcbi.1008013`. doi:10.1371/journal.pcbi.1008013.

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

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

[81] L. Beaulieu-Laroche, E. H. Toloza, M.-S. Van der Goes, M. Lafourcade, D. Barnagian, Z. M. Williams, E. N. Eskandar, M. P. Frosch, S. S. Cash, M. T. Harnett, Enhanced dendritic compartmentalization in human cortical neurons, Cell 175 (2018) 643–651.

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

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

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