PNAS RESEARCH ARTICLE | NEUROSCIENCE

Beyond spiking networks: The computational advantages of dendritic amplification and input segregation

Cristiano Capone[a,1,2,3], Cosimo Lupo[a,1], Paolo Muratore[b], and Pier Stanislao Paolucci[a]

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, San Diego, CA; received December 7, 2022; accepted October 11, 2023

The brain can efficiently learn a wide range of tasks, motivating the search for biologically inspired learning rules for improving current artificial intelligence technology. Most biological models are composed of point neurons and cannot achieve state-of-the-art performance in machine learning. Recent works have proposed that input segregation (neurons receive sensory information and higher-order feedback in segregated compartments), and nonlinear dendritic computation would support error backpropagation in biological neurons. However, these approaches require propagating errors with a fine spatiotemporal structure to all the neurons, which is unlikely to be feasible in a biological network. To relax this assumption, we suggest that bursts and dendritic input segregation provide a natural support for target-based learning, which propagates targets rather than errors. A coincidence mechanism between the basal and the apical compartments allows for generating high-frequency bursts of spikes. This architecture supports a burst-dependent learning rule, based on the comparison between the target bursting activity triggered by the teaching signal and the one caused by the recurrent connections, providing support for target-based learning. We show that this framework can be used to efficiently solve spatiotemporal tasks, such as context-dependent store and recall of three-dimensional trajectories, and navigation tasks. Finally, we suggest that this neuronal architecture naturally allows for orchestrating "hierarchical imitation learning", enabling the decomposition of challenging long-horizon decision-making tasks into simpler subtasks. We show a possible implementation of this in a two-level network, where the high network produces the contextual signal for the low network.

dendritic amplification | pyramidal neuron | target-based learning | hierarchical imitation learning

Significance

Biological networks embody the capability to flexibly support a disparate variety of functions, learning with astonishing energy and sample efficiency. This motivates the search for biologically inspired learning rules for improving artificial intelligence. We propose a generation of bio-inspired networks composed of three-compartment neurons (inspired by L5 pyramidal neurons) and a paradigm for learning and computing. This generation of networks achieves properties that are inaccessible to standard networks, among which the capability to learn without error propagation, the possibility to robustly perform context-dependent tasks, and the possibility to implement hierarchical policies decomposing complex long-horizon tasks. Finally, we argue that our model is capable of accounting for context-dependent perceptive facilitation observed in humans and animals.

Biological networks of neurons can solve a disparate variety of tasks with high energetic and sample efficiency, motivating the search for biologically inspired learning rules for improving artificial intelligence. The last decades have seen consistent progress in the development of efficient neural networks (Fig. 1*A*), taking more and more inspiration from biology. The first generation of neural networks was based on perceptrons (also referred to as McCulloch–Pitts neurons or threshold gates), only capable of providing a digital output (as discussed in ref. 1). The second generation was based on computational units that apply an "activation function" with a continuous set of possible output values to a weighted sum (or polynomial) of the inputs. Typical examples are feedforward and recurrent sigmoidal neural networks. In the 90s, experimental results from neurobiology led to a third generation of neural network models, employing spiking (or "integrate-and-fire") neurons as computational units (1–4). Networks of spiking neurons are, with regard to the number of neurons that are needed, computationally more powerful than these other neural network models (1). Moreover, they allow improved energy efficiency (5–7) and the possibility to encode information through spike timing. Despite these theoretical and technological advancements, most biologically inspired neural networks are composed, so far, of point neurons (2, 8), and cannot achieve state-of-the-art performance of artificial intelligence algorithms, e.g., they struggle to solve the credit-assignment problem (9).

Recent findings on dendritic computational properties (10) and on the complexity of pyramidal neurons dynamics (11) motivated the study of multicompartment neuron models in the development of biologically plausible learning rules (9, 12–14). These observations are giving rise to a fourth generation of neural networks, composed of spatially extended, bursting neurons, making it possible to exploit learning paradigms that were not accessible in the previous generations of neurons, as we will demonstrate in this paper.

It has already been proposed that segregation of dendritic input (13) (i.e., neurons receive sensory information and higher-order feedback in segregated compartments) and

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]C.C. and C.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: cristiano0capone@gmail.com.

[3]Present address: National Center for Radiation Protection and Computational Physics, Istituto Superiore di Sanità, Rome 00161, Italy.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2220743120/-/DCSupplemental.

Published November 29, 2023.

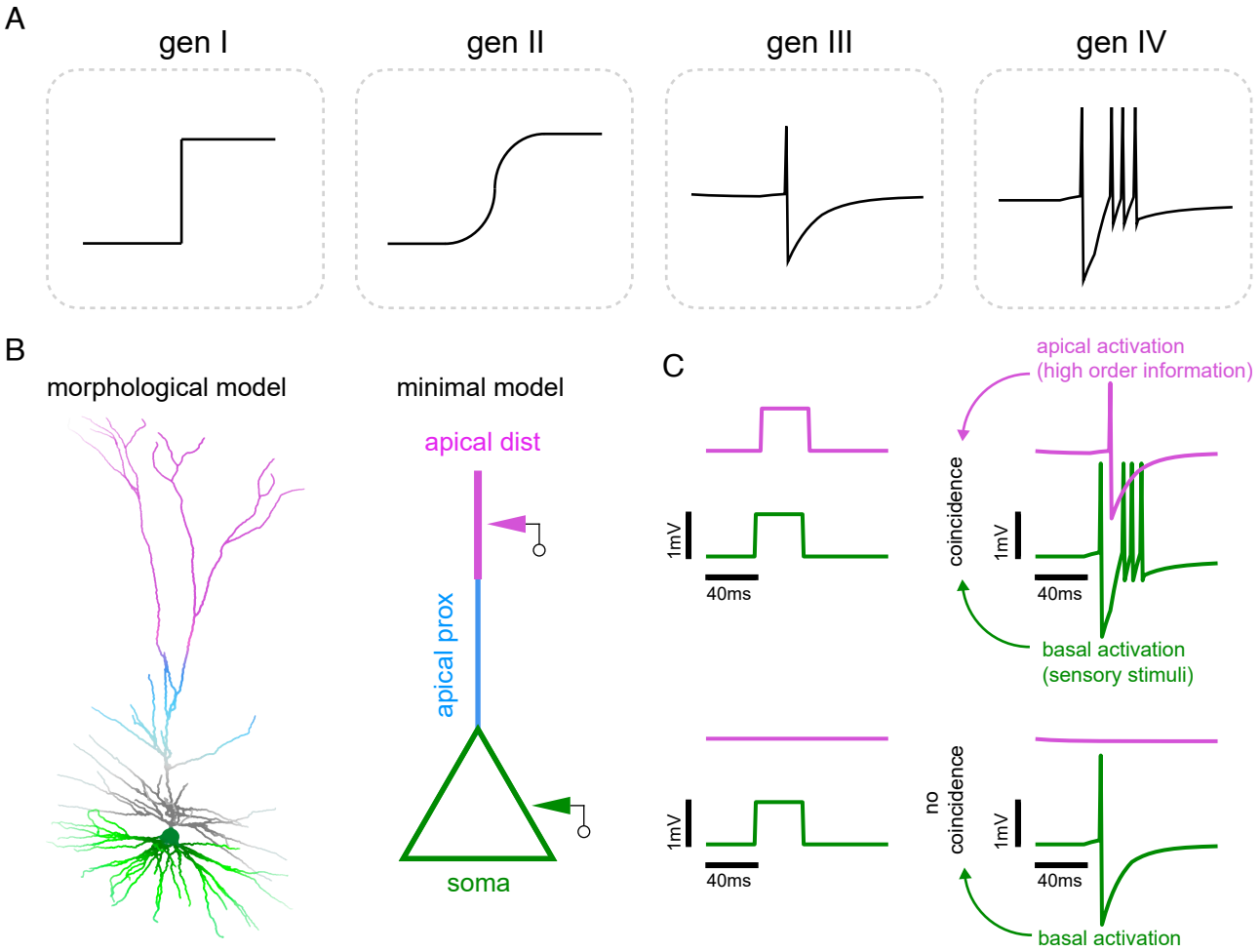

Fig. 1. The multicompartment neuron model. (*A*) The four generations of neural networks, from "threshold gate" and "activation function" models to spiking neurons and finally to multicompartment neurons producing high-frequency bursts. (*B*) (*Left*) Representation of the morphology of a pyramidal neuron. The neuron structure, depicted for representational purposes, is taken from the Allen Brain Atlas (19) and is a pyramidal neuron from the middle temporal gyrus of the human brain, left hemisphere. (*Right*) Our three-compartment simplified model of a L5 pyramidal neuron: The soma (green) receives sensory inputs; the apical proximal compartment (blue) receives the recurrent connections from the other neurons in the network; the apical distal compartment (purple) receives teaching/contextual signals from other areas of the cortex. (*C*) The coincidence mechanism implemented thanks to dendritic segregation, taking inspiration from the mechanism shown in ref. 11. (*Top*) When a dendritic spike occurs in coincidence with a somatic spike, a high-frequency burst of somatic spikes is generated. (*Bottom*) When this coincidence does not occur, only isolated spikes can be generated.

generation of high-frequency bursts of spikes (9) would support backpropagation in biological neurons. However, current approaches require propagating errors with a fine spatiotemporal structure to all the neurons, and it is not yet clear whether this is possible in biological networks.

For this reason, in the last few years, target-based approaches (3, 15–18) started to gain more and more interest. In a target-based learning framework, the targets—rather than the errors—are propagated through the network (15, 17). In this way, it is possible to directly suggest to the network the internal solution to a task (3, 4, 16). This approach requires the simultaneous evaluation of the spontaneous activity and the target activity of the network (3, 16), in order to adjust parameters and make the network converge to the target dynamics. This is usually implemented by creating two identical replicas of the network: a first one for evaluating the target dynamics and a second one in which parameters (such as recurrent weights) are optimized. This paradigm, however, lacks biological plausibility.

In the present work, we show that bursts and dendritic input segregation offer a natural solution to this dilemma. Our learning rule builds upon an important architectural assumption (Fig. 1*B*): The input arriving to the apical dendritic compartments is further segregated in local predictions (to the proximal apical compartment) and teaching/contextual signals (to the apical distal compartment). A coincidence mechanism between the basal and the apical distal (or the apical proximal) inputs generates a burst (11), eventually defining the target (or the current) spatiotemporal bursting dynamics of the network. In this way, target and current activities are simultaneously evaluated in the same neuron, and the presence of two replicas of the same network is no longer required.

This segregation, besides being a natural way to compare local predictions and higher-order suggestions, is supported by experimental tracking of afferent fibers (see e.g., Box 1 in ref. 11) and can be justified by geometric considerations: As the mentioned signals come from very different spatial locations, it is reasonable to assume that they would arrive in different regions of the neuron. Our assumption could also be interpreted as a theoretical prediction to be validated by dedicated experiments.

In our model, we exploit dendritic computation to let arbitrary signals act as teaching signals which drive the learning procedure in a biologically plausible fashion. This allows us to flexibly

store and recall arbitrary trajectories, with performances that are competitive with the state-of-the-art error-based approaches. Finally, we show that this neuronal architecture naturally allows for orchestrating hierarchical imitation learning, enabling the decomposition of challenging long-horizon decision-making tasks into simpler subtasks (20, 21), through the implementation of a two-level network, with the high network acting as a "manager" and producing the contextual signal for the low network, the "worker."

Results

Target-Based Learning with Bursts.

Neuronal architecture. Inspired by the morphology of L5 pyramidal neurons (Fig. 1 *B*, *Left*), we defined a neuron model composed of three separated compartments (Fig. 1 *B*, *Right*): the basal one (i.e., the soma), receiving the sensory input; and two apical ones, the proximal apical compartment, receiving recurrent connections from the somatic compartment of the other neurons in the network, and the distal apical compartment, receiving the context/teaching signal from other areas of the cortex, with a higher level of abstraction. We want to emphasize that our model describes the soma and basal dendrites as one somato-basal compartment, aiming to keep the model as simple as possible. Consequently, we often use the terms "somatic" and "basal" compartment interchangeably, as they both refer to the same compartment of our model. Here, the distinction lies solely in semantics: "basal" typically refers to input properties, while "somatic" relates to output properties such as spikes and bursts.

Each of these compartments is characterized by a membrane potential modeled through leaky-integrate-and-fire dynamics. The spike emitted by the soma of the i-th neuron is described by variable z_i^t, which is equal to 1 when the spike is emitted at time t and 0 otherwise. The spikes emitted by the proximal and distal apical compartments are then described by variables a_i^t and $a_i^{\star,t}$, respectively. The underlying idea is that the distal compartment provides a target for the proximal one, motivating the use of the superscript symbol $\star$, which indicates the variables concerning the target.

Following ref. 11, a coincidence mechanism between the basal and the apical compartments has been implemented, yielding high-frequency bursts of spikes from the soma. In more detail, after a somatic spike, $z_i^t = 1$, a coincidence window is opened for a time interval ΔT. This is described by the variable $\overline{z}_i^{t'}$, the indicator function for $t' \in [t, t+\Delta T]$, which is 1 during this time window and 0 elsewhere. If a spike is generated by the distal or proximal apical compartment within such time window, $a_i^{t'} = 1$ or $a_i^{\star,t'} = 1$, with $t' \in [t, t+\Delta T]$, a high-frequency burst of spikes is then produced (Fig. 1 *C*, *Top*); otherwise, the somatic spike remains isolated (Fig. 1 *C*, *Bottom*). Such a coincidence mechanism is defined for both the distal and the proximal apical compartments; the functional differentiation between them will be clarified in the following section. The resulting proximal and distal burst variables can be hence written respectively as:

$$\mathrm{B}_i^{t+1} = \overline{z}_i^t \, a_i^{t+1} \qquad [1]$$

$$\mathrm{B}_i^{\star,t+1} = \overline{z}_i^t \, a_i^{\star,t+1} \qquad [2]$$

Burst-mediated plasticity rule. The modular architecture of L5 pyramidal neurons, together with dendritic active computation, is thought to have astonishing computational properties (10). We propose that the multimodular architecture of our model can be exploited to develop alternative classes of learning rules. Starting from general considerations, we note how, because of the spatial segregation of the neuron, we can expect that a generic learning rule is a function of the activities of each compartment and can be described by the expression $f(z_i, a_i, a_i^\star)$. If we request space and time locality, then, we can further refine our description and impose a generic plasticity rule for the recurrent soma-to-proximal weights $w_{ij}^{s\to p}$ of the form: $\Delta w_{ij}^{s\to p} = f_{\text{post}}(z_i, a_i, a_i^\star)\, g_{\text{pre}}(\mathbf{s}_j)$, where f_{post} and g_{pre} are generic functions.

In this work, we present a particular instance of this class of burst-dependent plasticity rules, which naturally enables target-based learning. More specifically, we propose that the pattern of bursts defined by the proximal compartment (receiving the recurrent connections $w_{ij}^{s\to p}$ from the network, see Fig. 2*A*, blue arrow) should mimic the ones induced by the distal compartment (which receives the teaching signal; see Fig. 2*A*, purple triangle), with the sensory input entering the somatic compartment (see Fig. 2*A*, green triangle). This is made possible by using the following plasticity rule for recurrent weights $w_{ij}^{s\to p}$ (which can be derived analytically through a likelihood maximization, see *Methods* for details):

$$\Delta w_{ij}^{s\to p} = \eta \left[a_i^{\star,t+1} - a_i^{t+1} \right] \overline{z}_i^t \, e_j^t, \qquad [3]$$

where η is the related learning rate and $e_j^t = \partial u_i^t / \partial w_{ij}^{s\to p}$ is referred to in the literature as the spike response function (12), with u representing the membrane potential of the proximal apical compartment. In other words, such plasticity rule aims at aligning in time apical proximal spikes with apical distal ones when the somatic window $\overline{z}_i^t$ is open. We remark that such a learning rule can be computed online and only requires observables which are locally accessible to the synapses in space and time.

These ingredients allow to arbitrarily train a neuron to produce a burst by using the proper teaching signal (Fig. 2*A*). For example, let us consider a neuron that produces a somatic spike in response to a sensory input (see Fig. 2*B*, first column). To induce this neuron to produce a burst in response to that sensory stimulation, it is sufficient to inject a teaching current capable of inducing a spike in the apical distal compartment and hence a burst in the soma (see Fig. 2*B*, second column). The mismatch between the proximal and the distal response (see Fig. 2*B*, second column, blue and purple lines) triggers synaptic plasticity, inducing an increase of synaptic weights (see Fig. 2*B*, second column, blue dashed line). After training, the pyramidal neuron is capable of producing a burst (Fig. 2*B*, third column) also when the teaching signal is no longer present (see Fig. 2*B*, last column), thanks to the proximal apical spike induced by the learned recurrent weights.

An important feature of our model is what we call teacher neutrality, i.e., the presence of the teaching signal becomes irrelevant after the training (see Fig. 2*B*, comparison between third and last column). Indeed, if the apical proximal compartment emits a spike when the somatic window is already open, and a somatic burst is consequently generated, a further apical distal spike does not trigger more bursts. This feature is essential for target learning and can be biologically justified by the fact that the burst mechanism is a strongly nonlinear phenomenon, which depends on a threshold activation of the apical calcium channels. This means that it is either activated or not: If it is already activated by the recurrent weights, then, a further excitatory signal (the teaching signal) marginally affects bursting dynamics (11). This mechanism, to which we refer to as apical saturation, is properly reproduced by our model.

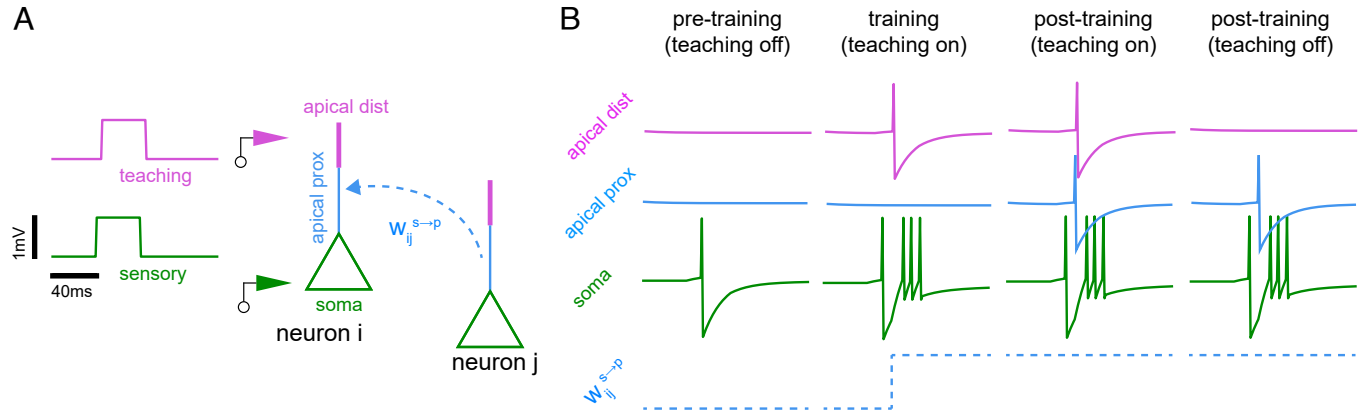

Fig. 2. Teaching through burst-mediated plasticity rule. (*A*) Our three-compartment neuron receives spatially segregated inputs: The somatic compartment (green) receives sensory inputs; the apical proximal compartment (blue) receives the recurrent connections from the somatic compartment of the other neurons in the network; the apical distal compartment (purple) receives teaching/contextual signals from other areas of the cortex. (*B*) Schematics for the teaching process. Time course of the membrane potentials of the three compartments (colors as in panel *A*) and of a synaptic weight (blue dashed). The plasticity rule we derived allows for synaptic changes only when a burst occurs. It aims at aligning proximal spikes (*a*) generated by the network, with distal spikes ($a^\star$) induced by the teacher/context. (First column) Before learning, recurrent connections are not trained, and no activity from the apical proximal compartment is detected. (Second column) The coincidence between a somatic spike (green) and an apical distal spike (purple) triggers a burst (green). The mismatch between distal (purple) and proximal spikes (blue) when the somatic coincidence window is open induces the change of presynaptic weights during the training phase (dashed blue). (Third column) After learning, the network is able to self-sustain the apical activity and hence correctly reproduce the target determined by the teaching signal. (Last column) The neuron is capable of producing the learned burst also when the teaching signal is no longer present.

Store and recall. We begin by demonstrating how our architecture naturally supports target-based learning. To do this, we focus on a standard benchmark: the store and recall of a 3D trajectory $y_k^{\star,t}$ ($k = 1, \dots, 3$; $t = 1, \dots, T$; T = 1,000). We consider a network of $N = 500$ neurons (see Fig. 3*A*, 400 bursting neurons with the pyramidal architecture described above, plus 100 nonbursting point neurons; more details in the *Methods*). We chose $y_k^{\star,t}$ as a temporal pattern composed of 3 independent continuous signals, each of which specified as the superposition of the four frequencies $f_n \in \{1, 2, 3, 5\}$ Hz, with uniformly extracted random phases $\phi \in [0, 2\pi]$:

$$y_k^{\star,t} = \sum_{n=1}^{4} A_{k,n} \cos\left(2\pi f_n t + \phi_{k,n}\right), \quad k = 1, 2, 3. \qquad [4]$$

Amplitudes $A \in [0.5, 2.0]$ are randomly extracted as well, eventually normalized in order to have trajectories within the $[-1, 1]$ interval. These target trajectories are randomly projected through a Gaussian matrix with variance σ^2_{targ} to the (distal) apical dendrites of the network as a teaching signal. This input shapes the spatiotemporal pattern of spikes $a_i^{\star,t}$ from the distal apical compartment as well as the related target spatiotemporal pattern of bursts $\mathrm{B}_i^{\star,t}$ (Fig. 3 *B*, *Bottom*, brown points) as described above.

A clock signal serving as a sensory input (Fig. 3*A*) is randomly projected (through a Gaussian matrix with variance σ^2_{in}) to the basal dendrites. In more detail, the clock is here modeled as a sort of time-step function with I steps, such that at each time t only component $i = \lfloor I \cdot t/T \rfloor$ is equal to 1, while others are 0 (see *SI Appendix*, Table S1 for model parameters).

Before learning and in the absence of the teaching signal, the network autonomously produces a spatiotemporal pattern of bursts, that does not encode any relevant information (see Fig. 3 *B*, *Bottom*, first column). The internal bursting activity is translated into the output y (see Fig. 3 *B*, *Top*, first column) by means of a read-out matrix w_{out}, randomly initialized and to be trained following the rule derived by minimizing the mean squared error (mse) between the target output and the network output:

$$\Delta w_{ki}^{\text{out}} = \eta_{\text{out}} \left[y_k^{\star,t} - \sum_h w_{kh}^{\text{out}} \hat{\mathrm{B}}_h^t \right] \hat{\mathrm{B}}_i^t, \qquad [5]$$

where η_{out} is the related learning rate (possibly different from that of recurrent connections) and $\hat{\mathrm{B}}$ is a time-filtered version of burst variable B (see *Methods* for details).

In order to train the network, the teaching signal given by the target trajectory $\boldsymbol{y}^\star$ is switched on; the network dynamics consequently change, producing a pattern of bursts (see Fig. 3 *B*, *Bottom*, second column) that is a superposition of the ones originally generated by the randomly initialized network and the ones induced by the teacher. The resulting mismatch between the network predictions (B_i^t) and the teacher suggestions ($\mathrm{B}_i^{\star,t}$) activates the plasticity of recurrent weights $w_{ij}^{s \to p}$, according to Eq. **3**. This mismatch is represented in Fig. 3*D* for different training epochs, by means of the rastergram of the internal local error $\hat{\mathrm{B}}_i^{\star,t} - \hat{\mathrm{B}}_i^t$. In this representation, red points are target bursts that are not produced autonomously (i.e., in the absence of the teacher) by the network, while blue ones are "undesired" bursts not present in the target. It is evident how the change in recurrent weights, induced by the training procedure, progressively reduces such mismatch. We stress that isolated somatic spikes are instead not explicitly constrained by our learning rule and hence are free to move during training.

At the end of learning, the network only produces the bursts suggested by the teacher (induced by the teaching signal; see Fig. 3*B*, third column), correctly encoding the solution of the task, i.e., the 3D trajectory, through the learning of readout weights as well (see Fig. 3 *B*, *Top*, third column). The network is eventually capable of reproducing the target spatiotemporal pattern of bursts and the target output in the absence of the teaching signal (see Fig. 3 *B*, *Top*, fourth column). The ability of performing the same dynamics equally in the presence or absence of the teaching signal is a further demonstration of the outstanding property of teacher neutrality illustrated above.

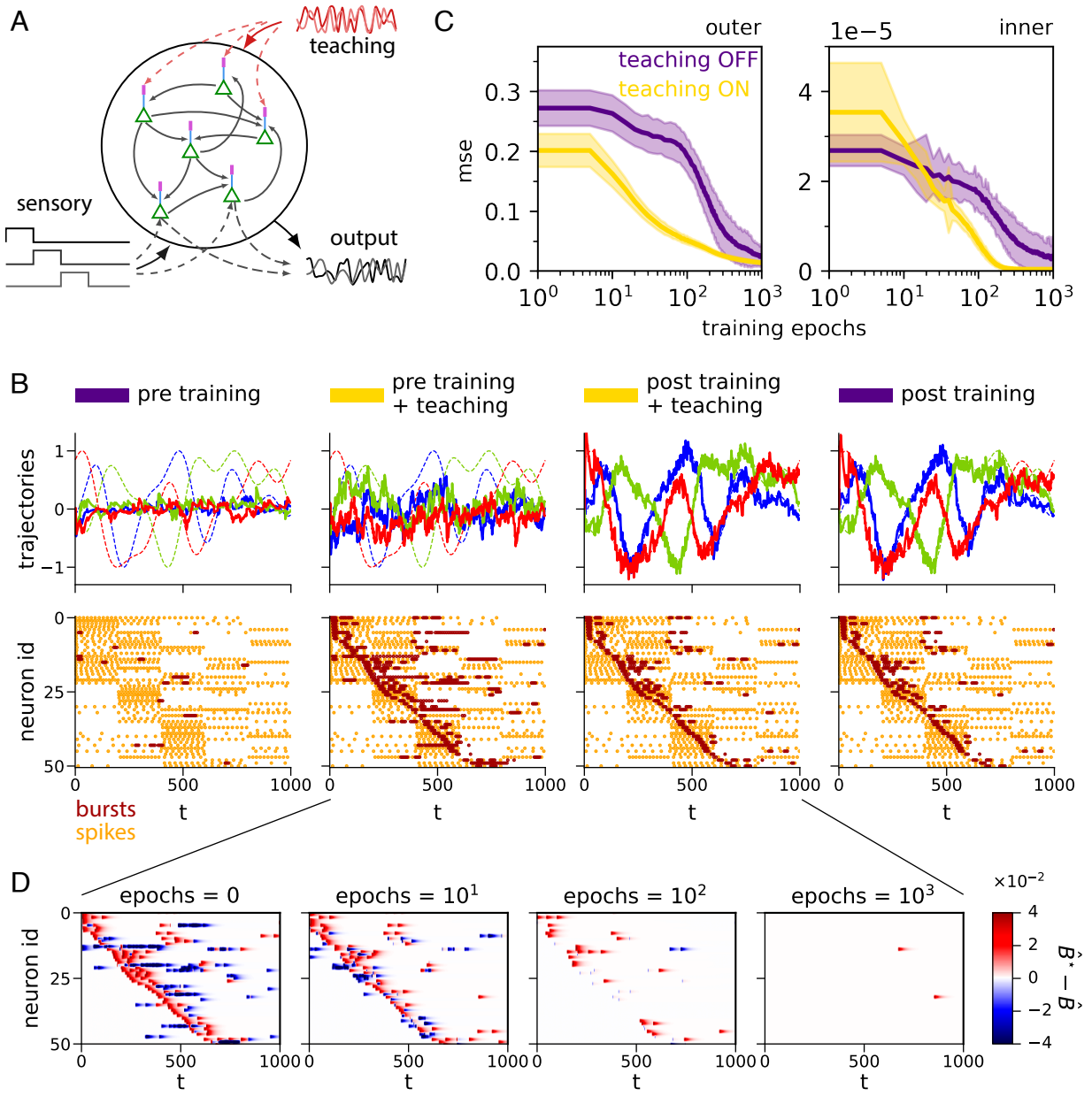

Fig. 3. Model structure. (*A*) Sketch of the network setting used for the store-and-recall task of a 3D trajectory. (*B*) For each of the four panels, in the *Top* row, we reported the target trajectory (dashed lines) together with the output produced by the network (solid lines); in the *Bottom* row, isolated spikes (yellow) and bursts (brown) from the somatic compartment. Before learning (first column), the network randomly produces patterns of bursts. When the teaching signal is present (second column), the network produces a combination of the initial bursts and of those induced by the teaching signal. After the training procedure (third column), the network learned to produce only the correct bursts. Eventually, the network is able to autonomously reproduce the proper spatiotemporal pattern of bursts and output also in the absence of the teaching signal (fourth column). In *Bottom-Row* panels, for visualization purposes, neurons are sorted according to the "post training + teaching" time of first burst, and then, only a subset of them is actually shown. (*C*) (*Left*) mse of the output 3D trajectory against the target one during the training phase with (yellow) and without (violet) the suggesting signal. (*Right*) mse between the internal pattern of bursts and the one suggested by the teaching signal, with (yellow) and without (violet) the suggesting signal. Averages over 50 independent realizations. (*D*) Rastergram of the internal local error $\hat{B}_i^{*,t} - \hat{B}_i^t$ for different training epochs, from the beginning to the end of the teaching procedure. We used the same neuron indexing as for *Bottom-Row* panels in *B*.

Target-based learning allows to define both an outer mse, evaluated between the target 3D trajectory and the one reproduced by the network, and an inner mse, evaluated between the target spatiotemporal pattern of burst and the one produced by the internal activity of the network (Fig. 3 *C*, *Left* and *Right* panel, respectively).

The final (outer) mse we get is approximately 0.01 (averaged over 50 realizations), to be compared with 0.01 obtained by ref. 2 and 0.001 obtained by (3). Aiming at a fair comparison, a relevant difference has to be pointed out between the present work and the aforementioned two: Here, we encode the target only through bursts, which are way less than spikes (used by refs. 2 and 3), resulting in a sparser and hence more noisy spatiotemporal encoding. This drawback is however counterbalanced by a remarkable improvement in terms of biological plausibility: i) a target-based instead of an error-based framework, which does not require the propagation of a precise spatiotemporal structure of the error (a feature that has not been observed

experimentally in biological networks of neurons), and ii) a natural definition of the target in terms of internal bursting activity, that does not require two replicas of the network (refs. 3 and 16).

Apical Contextual Signals to Robustly Select Desired Responses. The distal apical compartment is designed not only to receive teaching signals but also contextual information from other areas of the cortex, acting as a hint for the task to address. With this idea in mind, in this section, we show that it is possible to exploit different context signals (projected through a Gaussian random matrix with variance σ^2_{cont}) to flexibly select and recall one of the trajectories previously stored in the network. In the simplest configuration, two different contexts, A and B, can be modeled through 2D time-constant binary signals projected on the distal apical compartment, $\chi_{(1)} = (1, 0)$ for A and $\chi_{(2)} = (0, 1)$ for B (Fig. 4*A*). During the training, each context is associated with a well-defined target to learn (again a 3D trajectory, as defined in the previous section). In Fig. 4 *B*, *Left Side*, they are reported in red and black, respectively (only one of the three trajectories for each target is shown, for simplicity); the same color coding is used for associated context signals. To stabilize the learning, we exploited the trick of halving the learning rates η and η_{out} every 100 training iterations. The orthogonality of contexts and related targets is further stressed by imposing a sparsification (of 75% in the present case) in the random matrices we use to project the context and the target on the apical compartments of the network.

During the recall phase, the teacher signal is no longer present, while the context signal suggests the network which of the learned trajectory to reproduce. We show that when the context is projected to the network, the desired output is correctly recalled (Fig. 4 *B*, *Left Side*). Moreover, if the context signal is turned off in the middle of the trajectory, the network is still able to self-sustain its inner dynamics, thanks to recurrent connections (Fig. 4 *B*, *Left Side*), and correctly replicate also the remaining part of the selected trajectory. In Fig. 4*C*, we reported the mse (averaged over 10 realizations), measured against both the correct target trajectory (purple square markers) and the wrong one (purple round markers), i.e., the one associated with the other context signal, both before context turnoff and after it. In other words, the context works here as a "suggestion," so that once started the reproduction of the correct output trajectory, the context itself becomes useless.

To demonstrate the importance of projecting the context signal on the apical compartments, we compare these results with the case in which the context is projected on the basal ones (both during the training and the retrieval phases). As a consequence, the context is not integrated in a dedicated compartment, but it is summed together with sensory inputs (represented in this experiment by the clock signal). In this case, the desired trajectory is correctly retrieved only when the context is on (Fig. 4 *B*, *Right Side*), while after the turnoff, the network is no longer able to self-sustain bursts creation, in turn causing a dramatic drop in the retrieval performances (Fig. 4 *B*, *Right Side*). In other words, we observe that the basal context is now interpreted as a necessary input, and not as a suggestion for recalling the correct output. Corresponding average mse, again for both the correct target trajectory and the wrong one, both before turnoff and after it, is now reported in green in Fig. 4*C*.

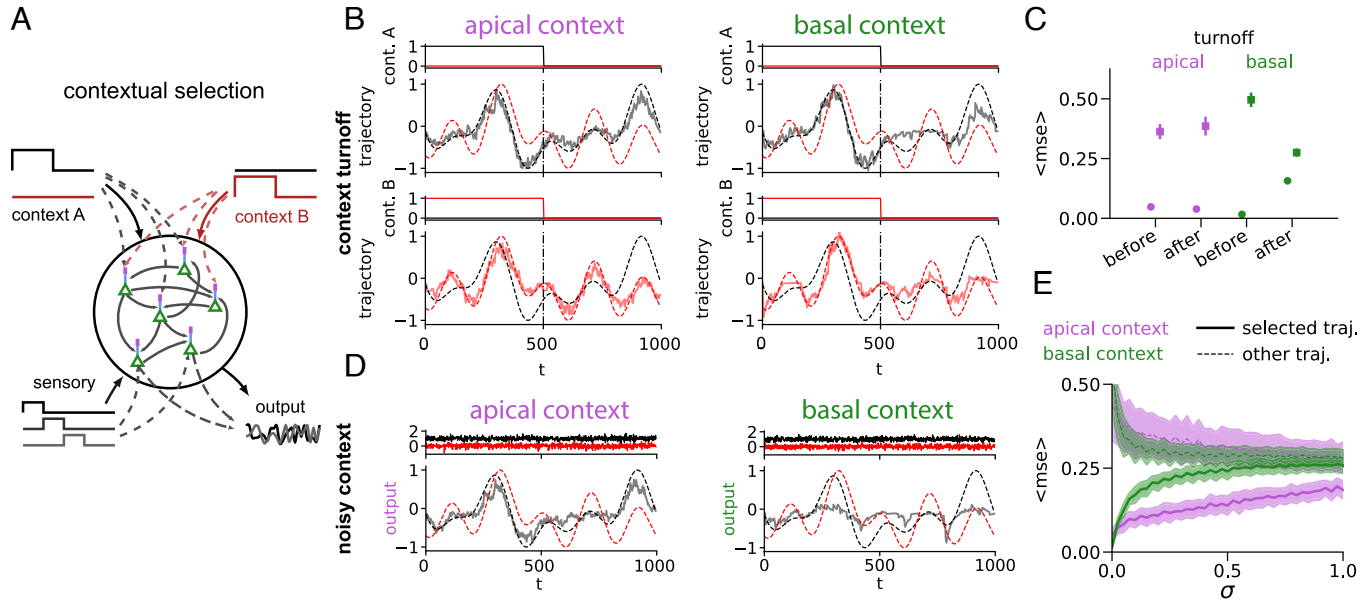

Fig. 4. Apical signals for contextual selection. (*A*) Sketch of a network of pyramidal neurons, where a binary context signal (*A* or *B*) is projected on the apical distal compartment. Given the same sensory input, the target output changes accordingly to the context. (*B*) (*Left*) The network is able to reproduce the correct output trajectory even if the context is provided only in the first time steps ("turnoff" experiment). Only one component of the 3D trajectory is represented for each context/target pair. (*Right*) An alternative model in which the context is projected on the basal compartment is no longer able to reproduce the correct output trajectory after the turnoff. (*C*) Summary of performances of the two model versions: context projected on apical (purple) vs basal (green) compartment during turnoff test in the middle of the trajectory. Mean squared error in the second part of the trajectory (no context) is compared with respect to error in the first part (context still active); mean and SD are intended over 10 independent network/target realizations (round markers). mse between the output and the trajectory selected by the other context is also reported, as a reference (square markers). (*D*) (*Left*) Even in the presence of a noisy context signal (*Top* inset, Gaussian white noise with $\sigma = 0.2$), the desired trajectory *A* (black) is reliably reproduced in output. (*Right*) When using a noisy basal context (*Top*, *Inset*, $\sigma = 0.2$), at variance, the desired target *A* can no longer be reproduced by the network. (*E*) Average performances of the apical/basal context (purple/green, respectively) as a function of the noise SD σ. Solid lines: mse between the output and the selected target trajectory. Dashed lines: mse between the output and the trajectory selected by the other context. Averages and error bars are intended over 10 independent network/target realizations.

Furthermore, we observed that our architecture exhibits robustness against corruption in the context signal, which may be the case when at higher cortical levels, there is only a mild preference in favor of which strategy to adopt (in comparison with the training phase, where each target is clearly and unequivocally associated with a sharp context signal). Here, a Gaussian white noise with zero mean and variance σ^2 is added during the test to context signals exploited in the training (Fig. 4 *D*, *Left*, for $\sigma = 0.2$). The produced trajectory is very similar to the trajectory selected by context A (black dashed line) and different from the trajectory selected by context B (red dashed line). Fig. 4*E* shows the mse (averaged over 10 independent realizations of the experiment) between the output and the target trajectory (solid purple line) as a function of σ. As a reference, we also report the mse between the output and the trajectory selected by the other context signal (dashed purple line). We observe a remarkable resilience of the network with apical context, while the network with basal context suddenly loses the ability to reproduce the desired output already at low levels of noise (Fig. 4 *D*, *Right*, for $\sigma = 0.2$ and Fig. 4*E*, green lines). At higher levels of noise, the basal context becomes in practice useless, while the apical one is still able to reproduce the target trajectory with a remarkably small error (Fig. 4*E*). Finally, we demonstrate that a standard spiking network exhibits performances comparable to the case where context is projected to the basal compartment (i.e., without a dedicated compartment, see *SI Appendix*), further confirming the superiority of our architecture.

This simple setup demonstrates how network processing of a sensory input can be strongly influenced by contextual signals and is coherent with context-induced perceptive distortions observed in humans and animals (11). For example, the same sound can bring different word perceptions, when reading different texts or looking at different lip movements. Similarly, in ref. 11, it is shown how the capability of a monkey to distinguish a figure from the ground was dependent on feedback connections to V1 (22).

Hierarchical Imitation Learning. The presence of an apical context that acts as a gating signal by flexibly selecting which dynamics to reproduce (and when) can be used as a building block for neural architectures that offer a biologically plausible implementation of hierarchical imitation learning (HIL). In this work, we propose a two-level hierarchical network where the higher submodule (high network or "manager") computes the optimal strategy and exploits the context signal as a communication channel with the lower submodule (low network or "worker"), which executes the selected task (Fig. 5*A*). We show that this architecture can efficiently solve the button & food task defined in a previous work of ours (4). In this scenario, an agent starts at the center of a square domain, in which are also present a button and an initially locked target (the "food;" see sketch of the task in Fig. 5*B*). The agent must observe the environment, make motor decisions to reach the button and unlock the food, and subsequently reach the food itself. In each episode, the button and the food are uniformly extracted on a unitary circle centered in the origin and in the button position, respectively. In spirit, this task is very similar to those also known as "door & key," "key-to-door," or "key & room" (23, 24). The global task effectively decomposes into two simpler subtasks or goals: reach_button and reach_food. The high network computes which of these two strategies (and when) to pursue and communicates it to the low network, which in turn implements the fine motor controls. Both the high and the low network share the same input ($I = 80$ input units), the horizontal and vertical differences of both button and food positions with respect to the agent location, $\Delta^t \equiv \{\Delta x_b^t, \Delta y_b^t, \Delta x_f^t, \Delta y_f^t\}$. Each of the Δ_i values is encoded by 20 input units with different Gaussian activation functions.

In this architecture, learning is implemented via a natural hierarchical extension of behavioral cloning (BC), a machine learning approach where a model learns to imitate the behavior of an expert by observing its actions. Here, an expert system provides a collection of hierarchical demonstrations $\mathcal{D} = (\mathcal{D}_{\mathrm{L}}, \mathcal{D}_{\mathrm{H}})$ for both submodules. A demonstration $\mathcal{D}_{\mathrm{L,H}}$ is a trajectory specification that can be described by the following tuple:

$$\mathcal{D}_{\mathrm{L,H}} = \left(\mathrm{state}_{\mathrm{L,H}}^t, \mathrm{action}_{\mathrm{L,H}}^t, \mathrm{goal}_{\mathrm{L}}^t\right), \quad t \in \{1, \ldots, T\}, \qquad [6]$$

where the goal^t component is missing for the high network as it sits at the top of the hierarchy. The $\mathrm{state}_{\mathrm{L,H}}^t$ component is the same for the two submodules and is equal to Δ^t, representing a description of the agent position in the environment. The high-network output $\mathrm{action}_{\mathrm{H}}^t$ encodes the selected strategy, and it is projected to the low network as a contextual signal to the distal apical compartment, where it is interpreted as the low network $\mathrm{goal}_{\mathrm{L}}^t$. We define this target behavior to be the 2D signal:

$$\boldsymbol{y}_{\mathrm{H}}^{\star,t} \equiv \boldsymbol{\chi}_{(1)}\Theta\left(t < t_{\blacksquare\!\cap}\right) + \boldsymbol{\chi}_{(2)}\Theta\left(t > t_{\blacksquare\!\cap}\right), \qquad [7]$$

where $\boldsymbol{\chi}_{(1)} = (1, 0)$ and $\boldsymbol{\chi}_{(2)} = (0, 1)$, and $t_{\blacksquare\!\cap}$ is the time when the button is reached (Fig. 5 *C*, *Top*). Intuitively, it selects the reach_button subpolicy for the first part of the task and then switches to reach_target, once the latter has been unlocked. Given the input $\mathrm{state}_{\mathrm{L}}^t$ and the context $\mathrm{goal}_{\mathrm{L}}^t$, the low network is tasked to reproduce as output $\mathrm{action}_{\mathrm{L}}^t$ the velocity vector $\boldsymbol{y}_{\mathrm{L}}^{\star,t} \equiv \boldsymbol{v}^t = (v_x^t, v_y^t)$, where velocity components are computed so to reach the selected target in a straight line with a constant speed (Fig. 5 *C*, *Center*). Both high- and low-network outputs are computed as linear read-outs of their internal bursting activities, as described for the store-and-recall task (Fig. 5 *C*, *Bottom*, for the low network).

We implement the cloning procedure in a supervised fashion, following the same procedure described for previous tasks: The two submodules are trained to reproduce their target outputs given their set of inputs (context included), exploiting the plasticity of read-out weights for both networks and of recurrent weights for the high network.

Finally, the two-level network is tested in closed loop in the environment described above, featuring the button and the food. Performances are measured via the following quantity:

$$\rho \equiv \Xi_{\blacksquare\!\cap}\,\frac{r_0}{\min_{t > t_{\blacksquare\!\cap}} d\left(\boldsymbol{x}_{\mathrm{agent}}^t, \boldsymbol{x}_{\mathrm{food}}\right)}, \qquad [8]$$

where $\Xi_{\blacksquare\!\cap}$ is the button-state indicator variable (0 when the button is locked and 1 otherwise), r_0 is the button and food size, and finally $d\left(\boldsymbol{x}_{\mathrm{agent}}^t, \boldsymbol{x}_{\mathrm{food}}\right)$ is the Euclidean distance between the agent and food positions at time t. The condition for a successful button-press (a switch from locked to unlocked state) and target-reach is taken to be $d\left(\boldsymbol{x}_{\mathrm{agent}}^t, \boldsymbol{x}_{\mathrm{btn/food}}\right) \leq r_0$, respectively. Note how this choice effectively prevents the apparent divergence in the expression for ρ as the episode is stopped when the target is reached, thus inducing a theoretical maximum achievable score of $\rho_{\max} = 1$. Otherwise, $0 < \rho < 1$ if the button has been

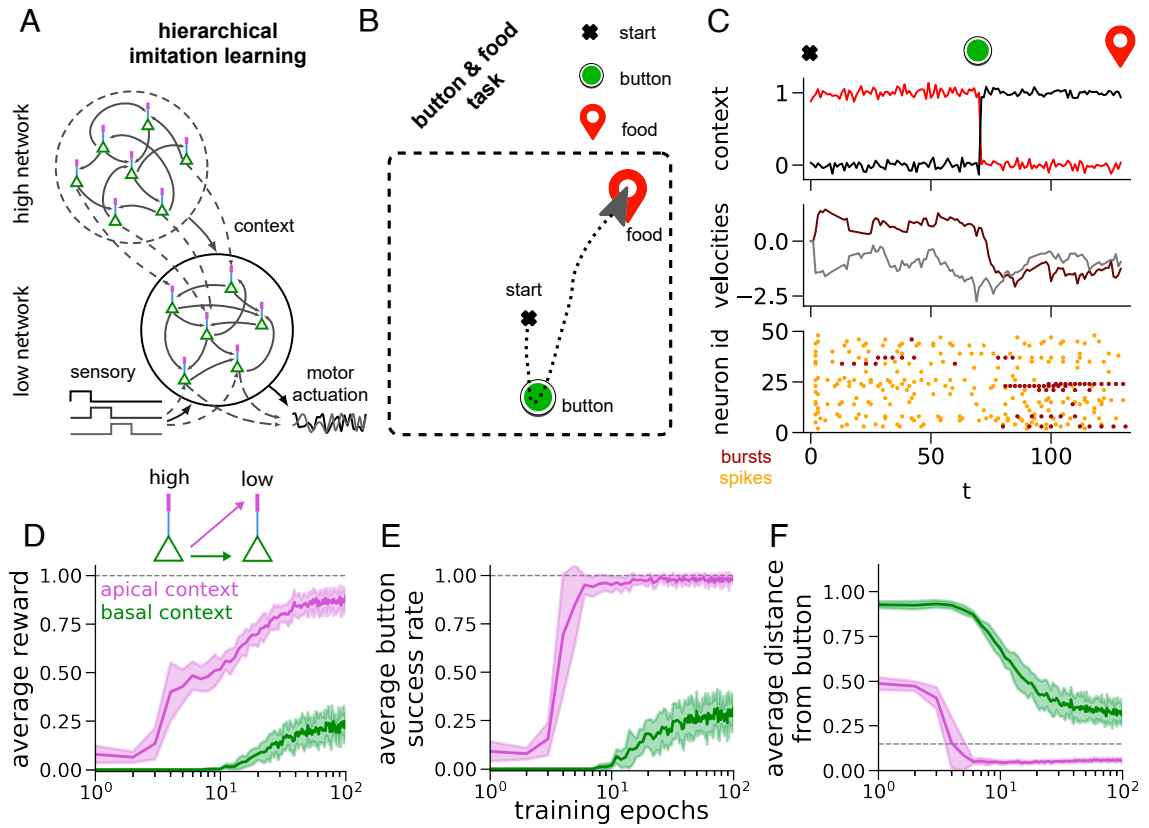

Fig. 5. Hierarchical imitation learning. (*A*) A two-level network, where high-level neurons produce a signal that serves as a context for the neurons in the low-level network, allows implementing hierarchical policies. The two subnetworks receive two different teaching signals in the training phase, in order to mimic a hierarchical policy from an expert demonstration. (*B*) A graphical sketch of the button & food task, where an agent placed at an initial position (black cross) in a 2D environment has to reach a button (green circle) in first place, so to unlock the food (red tag) and then reach for it. (*C*) The two-level network from panel *A* solving the button & food task presented in panel *B*. The high-level network chooses the order of the two subtasks (reach_button and reach_food) and when to switch from one to the other. It projects the instructions as a contextual signal (*Top Panel*) to the apical compartments of the low-level network. The low-level network then produces the output (velocities of the agent, *Center Panel*) necessary to solve the subtask as a read-out of its internal bursting activity (*Bottom Panel*, brown dots; orange dots represent the spiking activity). (*D*) Reward from closed-loop tests as a function of training epochs (average and SE over 10 network realizations, in lines and shadings respectively). Purple and green colors refer to the two different choices for the context projection on the low network: apical or basal compartments, respectively (see also *Inset* for a sketch of the two alternative models). The gray dashed line at 1.0 indicates the maximum possible reward achievable. (*E*) Average success rate for pushing the button as a function of training epochs. 1.0 is again the maximum possible value. Same color coding as in (*D*). (*F*) Average distance from the button at the end of the episode. The gray dashed line represents the button size. Same color coding as in (*D*).

unlocked, but the food has not been reached within the assigned time window T, or $\rho = 0$ if the button has not been reached at all.

After the presentation of 120 randomly positioned button-food pairs, we observe that the hierarchical two-level network learns to correctly and efficiently solve the button & food task, with an average final score $\rho = 0.88 \pm 0.04$ and over 70% of success rate (i.e., both button-press and target-reach conditions were met). In Fig. 5*D*, purple line, we report the average reward (over 10 independent network realizations) as a function of training epochs. Similarly, purple curves in Fig. 5 *E* and *F* show the average success rate in pushing the button, and the average minimum distance from the button, respectively.

We then run an additional experiment, where the high-network output is projected to the basal compartment of the low network rather than to the apical one, in the same spirit of the experiment performed above for the store-and-recall of 3D trajectories (Fig. 5 *D*, *Top Inset*). The results (averaged over 10 further independent realizations) are reported in Fig. 5 *D–F*, green lines. This choice leads to poor performances of the hierarchical policy ($\rho = 0.24 \pm 0.07$), demonstrating, also in this case, the necessity of a contextual signal of a different nature with respect to somatic input signals.

Hierarchical imitation learning in a spiking network. As a further validation of the computational advantages of input segregation, we want to compare the capability of the low network to integrate sensory and contextual signals demonstrated above with that of a network that does not have a dedicated compartment for contextual signals, e.g., a standard recurrent spiking network. To this aim, we still imagine this task to be solved by a two-level spiking network (Fig. 6*A*); though, for a completely fair comparison, here, the contextual signal provided by the high network is "exact," as given by an oracle, and we only consider the performances of the low network.

Following refs. 3 and 4, we consider an alternative version of the low network composed of single-compartment leaky-integrate-and-fire neurons (see *Methods* and *SI Appendix* for further details). In this case, the motor actuation described in the previous paragraph is decoded as a linear read-out from spiking activity z_i^t, rather than from bursts, where read-out weights are trained following the behavioral cloning procedure, i.e., by minimizing the mse between the behavior produced by the expert and the one produced by the network. Without compartmentalization, the neural input signal coming from the high network is summed to the inputs the neuron is

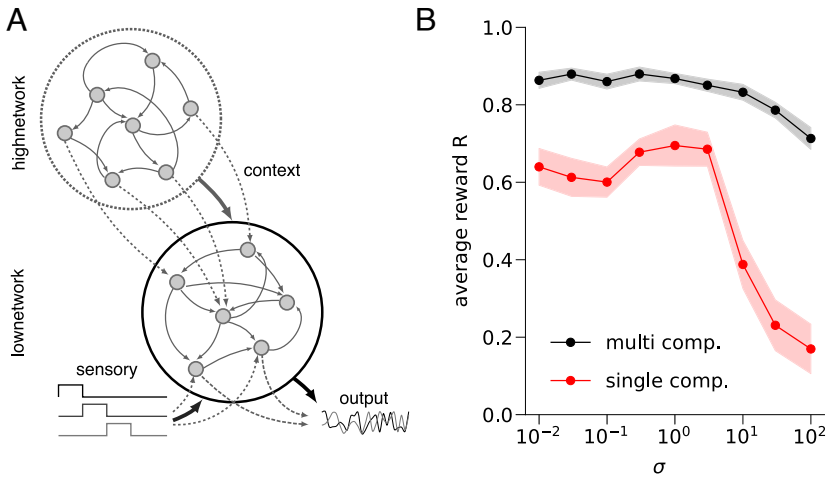

Fig. 6. Hierarchical imitation learning in a standard spiking network. (*A*) A two-level network with single-compartment neurons, where high-level neurons produce a signal that serves as a context for the neurons in the low-level network. In this case, the context is processed by low-level neurons by adding it to the other inputs, since the neurons have only one compartment. (*B*) Average reward on the button and food task (over 10 independent realizations) as a function of the amplitude of contextual noise, for the network composed of multicompartment (black curve) and single-compartment (red curve) neurons.

receiving from its recurrent connections and from sensory stimuli (encoding the environment information).

When comparing this single-compartment standard spiking architecture for the low network with the multicompartment bursting architecture presented before, we observe better performances for the latter choice ($\rho = 0.86 \pm 0.02$, averaged over 10 independent network realizations) compared to the former choice ($\rho = 0.67 \pm 0.03$). In addition, we also tested the robustness of the model to contextual noise, modeled as a Ornstein–Uhlenbeck process with zero mean, variance σ^2 and correlation time $\tau = 1.1$ ms added to the contextual signal provided to the low network. We observe superior stability in the bursting network (Fig. 6*B*, black curve), which is almost unaffected by contextual noise. On the other hand, the spiking network performances improve for small amounts of noise (which facilitate space exploration) and then dramatically drop for higher levels of noise (Fig. 6*B*, red curve).

We stress that it is not the explicit hierarchical implementation of the policy to improve the performance in the bursting network since it is also present for the spiking network. It is rather the lack of input segregation that causes the network to poorly perform with the task, reducing its robustness to contextual noise and hence making it more susceptible to little changes in the encoding of the subtask policy.

Methods

The Model. We defined a neuron model, inspired by the L5 pyramidal neuron, composed of three different compartments: a somatic (or basal) one (s) and two apical ones, named proximal (p) and distal (d), respectively (see Fig. 1*B* for reference). All the model parameters are reported in *SI Appendix*, Table S1.

If we consider the i-th of these multicompartment neurons, with $i = 1, \ldots, N_{\text{pyr}}$, its membrane potential vector $\mathbf{v}_i^t = \left(v_i^t,\, u_i^t,\, u_i^{\star,t}\right)$ (the membrane potentials of basal, proximal apical, and distal apical compartments, respectively) follows a leaky-integrate-and-fire dynamics with a characteristic time decay constant τ_m, which we can generically write as:

$$\mathbf{v}_i^{t+1} = \left[\left(1 - \frac{dt}{\tau_m}\right)\mathbf{v}_i^t + \frac{dt}{\tau_m}\mathbf{I}_i^{t+1}\right]\left(1 - \mathbf{s}_i^t\right) + \mathbf{v}^{\circlearrowleft}\mathbf{s}_i^t, \qquad [9]$$

where the vector quantities $\mathbf{I}_i^t = \left(I_{(s),i}^t,\, I_{(p),i}^t,\, I_{(d),i}^t\right)$, $\mathbf{s}_i^t = \left(z_i^t,\, a_i^t,\, a_i^{\star,t}\right)$ and $\mathbf{v}^{\circlearrowleft} = \left(\mathrm{v}_{(s)}^{\circlearrowleft},\, \mathrm{v}_{(p)}^{\circlearrowleft},\, \mathrm{v}_{(d)}^{\circlearrowleft}\right)$ represent the input current, the neuron spike and the reset potential, respectively, for each compartment (see the following sections for their explicit definitions). The membrane potential vector $\mathbf{v}_i^t$ defines the stochastic emission of the spike at the subsequent time step $\mathbf{s}_i^{t+1}$ via a sigmoid probability:

$$p\left(\mathbf{s}_i^{t+1}|\mathbf{v}_i^t\right) = \frac{\exp\left[\mathbf{s}_i^{t+1}\left(\frac{\mathbf{v}_i^t - \boldsymbol{v}_{\text{thr}}}{\delta v}\right)\right]}{1 + \exp\left(\frac{\mathbf{v}_i^t - \boldsymbol{v}_{\text{thr}}}{\delta v}\right)}, \qquad [10]$$

with $\boldsymbol{v}_{\text{thr}}$ being the firing threshold for the membrane potential of the three compartments (set to the same value $v_{\text{thr}} = 0$) and δv a model parameter controlling the probabilistic nature of the firing process (both assuming the same value for the three compartments, in our model). In the $\delta v \to 0$ limit, the spike-generation rule Eq. **10** becomes deterministic, with the sigmoid approaching the Heaviside function Θ:

$$p(\mathbf{s}^{t+1}|\mathbf{v}^t) = \Theta[\mathbf{s}^{t+1}(\mathbf{v}^t - \boldsymbol{v}_{\text{thr}})]. \qquad [11]$$

We remark that in all the numeric implementation of model dynamics, we assume the deterministic dynamics ($\delta v \to 0$).

Network composition. The network of N neurons is composed of N_{pyr} pyramidal bursting neurons, plus N_{pt} single-compartment nonbursting neurons, whose spiking dynamics is perfectly equivalent to that of the somatic compartment of pyramidal neurons (see below). While somatic input is projected to all the N neurons, network read-out is based on the bursting activity of pyramidal neurons only. Including these two types of neurons is a minimal way to account for the heterogeneity of the cortex composition, in the spirit of biological plausibility.

Bursting mechanism. We introduce the exponential filtering function filter (ξ^t, τ), defined recursively as:

$$\text{filter}\left(\xi^{t+1}, \tau\right) \equiv e^{-dt/\tau}\,\text{filter}\left(\xi^t, \tau\right) + \left(1 - e^{-dt/\tau}\right)\xi^{t+1}, \qquad [12]$$

starting from the initial condition $\text{filter}\left(\xi^0, \tau\right) = 0$. Basal spike signals are time-filtered through suitable time constants,

depending on the direction they propagate. Using the previous definition, we introduce the following time-filtered quantities:

$$\hat{z}_i^{t+1} \equiv \text{filter}\left(z_i^{t+1}, \tau_{\text{syn}}\right), \quad [13]$$

$$\hat{z}_{\text{ro},i}^{t+1} \equiv \text{filter}\left(z_i^{t+1}, \tau_{\text{ro}}\right), \quad [14]$$

$$\hat{z}_{\text{soma},i}^{t+1} \equiv \text{filter}\left(z_i^{t+1}, \tau_{\text{targ}}\right). \quad [15]$$

Such filtering is also applied to the adaptation term ω_i^t contributing to the input current of the basal compartment, which is time-smoothed as:

$$\omega_i^{t+1} \equiv \text{filter}\left(z_i^{t+1}, \tau_{\omega}\right). \quad [16]$$

The occurrence of a somatic spike opens a temporal somatic window:

$$\overline{z}_i^t = \Theta[\hat{z}_{\text{soma},i}^t - \vartheta_{\text{soma}}]. \quad [17]$$

Then, the onset of a burst in the proximal or distal compartments is induced by the coincidence between this somatic window $\overline{z}_i^t$ and an apical proximal spike a_i^t or an apical distal spike $a_i^{\star,t}$. Bursts can then be expressed, respectively, as:

$$\text{B}_i^{t+1} = \overline{z}_i^t a_i^{t+1}, \quad [18]$$

$$\text{B}_i^{\star,t+1} = \overline{z}_i^t a_i^{\star,t+1}. \quad [19]$$

Aiming for a time-window variable that is active during burst activity, we can iterate the same construction developed for spikes and consider the time-filtered burst-onset $\hat{\mathbf{B}}^t$:

$$\hat{\text{B}}_i^{t+1} \equiv \text{filter}\left(\text{B}_i^{t+1}, \tau_{\text{targ}}\right), \quad [20]$$

$$\hat{\text{B}}_i^{\star,t+1} \equiv \text{filter}\left(\text{B}_i^{\star,t+1}, \tau_{\text{targ}}\right). \quad [21]$$

One can again use these time-filtered quantities to introduce proximal and distal burst windows as:

$$\overline{\text{B}}_i^{t+1} \equiv \Theta[\hat{\text{B}}_i^{t+1} - \vartheta_{\text{burst}}], \quad [22]$$

$$\overline{\text{B}}_i^{\star,t+1} \equiv \Theta[\hat{\text{B}}_i^{\star,t+1} - \vartheta_{\text{burst}}]. \quad [23]$$

When at least one among proximal and distal bursts is above threshold, we finally have a neural burst activity window:

$$\overline{\text{B}}_{\vee,i}^{t+1} = \overline{\text{B}}_i^{t+1} \vee \overline{\text{B}}_i^{\star,t+1}, \quad [24]$$

which is the quantity that will appear in the dynamics of the compartments. The OR logical operator $\vee$ here enforces the fact that a coincidence can be generated when an apical signal is present, regardless of the specific apical compartment it comes from.

Basal/somatic compartment. The membrane potential of the basal compartment v_i^t evolves following Eq. **9**, with the following input contributions:

$$I_{(s),i}^t = \underbrace{\sum_{j=1}^{N} w_{ij}^{s\to s}\hat{z}_j^t}_{\substack{\text{recurrent soma-to-soma}\\ \text{connections}}} + \underbrace{\sum_{k=1}^{n_{\text{in}}} w_{ik}^{\text{in}} I_k^{\text{in},t}}_{\substack{\text{sensory}\\ \text{input}}} + \underbrace{\beta\overline{\text{B}}_{\vee,i}^t}_{\substack{\text{extra current}\\ \text{from coincidence}}} - \underbrace{b\,\hat{\omega}_i^t}_{\substack{\text{adaptation}\\ \text{term}}} + v_0. \quad [25]$$

In fact, recurrent connections targeting the basal compartments ($w_{ij}^{s\to s}$) are set to zero in our model. The contribution of sensory input is given by the input current $I_k^{\text{in},t}$, randomly projected to the neurons through the weights w_{ik}^{in}, while v_0 is a compartment-specific constant input. Finally, we introduce the extra current term $\beta\overline{\text{B}}_{\vee,i}^t$ (we set $\beta = 20$) to induce a high-frequency burst during the burst window $\overline{\text{B}}_{\vee,i}^t$, and the term $b\,\hat{\omega}_i^t$ taking into account spiking adaptation. Also, the basal reset potential is suitably increased during the burst window:

$$\mathbf{v}_{(s)}^{\circlearrowleft} = \frac{v_{\text{reset},s}}{1 + \alpha\overline{\text{B}}_{\vee,i}^t}, \quad [26]$$

where $v_{\text{reset},s} < 0$ is a compartment-specific scalar and α is a constant model parameter (we set $\alpha = 2$). Here, we employ the term "current" despite the fact that this quantity is measured in mV within the model since we set the membrane conductance $g_l = 1$ (as commonly adopted; see e.g., ref. 25, where the membrane capacitance is set to 1).

Apical proximal compartment. The apical proximal compartment of each neuron is connected to basal compartments of all the neurons through recurrent connections $w_{ij}^{s\to p}$. These recurrent connections are the object of the training procedure and are adjusted to reproduce the desired target. The dynamics of the membrane potential u_i^t follows Eq. **9**, with the input current defined as:

$$I_{(p),i}^t = \underbrace{\sum_{j=1}^{N} w_{ij}^{s\to p}\hat{z}_j^t(t)}_{\substack{\text{recurrent soma-to-proximal}\\ \text{connections}}} + u_0. \quad [27]$$

The reset potential for the proximal apical compartment $\text{v}_{(p)}^{\circlearrowleft} = v_{\text{reset},p}$ is a compartment-specific scalar, independent of burst activity, while u_0 is the compartment-specific constant input.

Apical distal compartment. The signal to be learned (the target) is considered as an input for the apical distal compartment: During the learning stage, it is injected via a projection matrix w_{ik}^{targ}, while it is completely absent during spontaneous activity. The coefficient $f_{\text{apic}} \in \{0, 1\}$ is used to gate this stage transition. Also, the input from the context (again randomly projected on the N neurons via the w_{ik}^{cont} matrix) is given as input for the apical distal compartment. So the total input current for the apical distal compartment is defined as:

$$I_{(d),i}^t = \underbrace{f_{\text{apic}}\sum_{k=1}^{n_{\text{output}}} w_{ik}^{\text{targ}} y_k^{\star,t}}_{\text{target/teach input}} + \underbrace{\sum_{k=1}^{n_{\text{cont}}} w_{ik}^{\text{cont}} C_k^t}_{\text{context input}} + u_0^{\star}, \quad [28]$$

where $y_k^{\star,t}$ is the target signal and C_k^t is the context signal, while $u_0^{\star}$ is the compartment-specific constant input. Also in this case, membrane potential $u_i^{\star,t}$ evolves according to Eq. **9**.

Derivation of the Learning Rule. We formulate the learning process as the maximization of the probability of observing the desired spatiotemporal pattern of bursts. By expressing such probability in terms of the recurrent connections in the network, we obtain an explicit expression for the learning rule. In doing so,

we directly extend previous approaches used for learning a target pattern of spikes (3, 26–28). We start by writing the probability to produce a burst in the i-th neuron at time t, given the somatic window $\bar{z}_i^t$. We propose the following compact formulation:

$$p(\mathrm{B}_i^{\star,t+1}|\bar{z}_i^t) = \frac{\exp\left[\mathrm{B}_i^{\star,t+1}\Phi_i^t(\bar{z}_i^t)\right]}{1+\exp\left[\Phi_i^t(\bar{z}_i^t)\right]}, \qquad [29]$$

where we have introduced $\Phi_i^t(\bar{z}_i^t) = a_i^{t+1}\bar{z}_i^t/\delta v - (1-\bar{z}_i^t)\gamma$. By definition, a burst can only happen by means of a basal-apical spike coincidence, represented by the $a_i^{t+1}\bar{z}_i^t$ term; when the basal window is open ($\bar{z}_i^t = 1$), the burst probability reduces to the usual sigmoidal function. When the window is closed and $\bar{z}_i^t = 0$, we have $\Phi_i^t(\bar{z}_i^t) = -\gamma$; we can thus tune the γ parameter to model the burst probability. In practice, we work in the $\gamma \to \infty$ limit where $\lim_{\gamma\to\infty} p(\mathrm{B}_i^{\star,t+1}|\bar{z}_i^t = 0) = 0$, which agrees with the intuitive understanding that a closed basal window prevents any burst activity. We introduce the log-likelihood $\mathcal{L}$ of observing a target burst activity $\mathbf{B}^\star$ given the soma-to-proximal connections $w_{ij}^{s\to p}$ as:

$$\mathcal{L}(\mathbf{B}^\star|\boldsymbol{w}^{s\to p}) = \sum_{i,t}\left[\mathrm{B}_i^{\star,t+1}\Phi_i^t(\bar{z}_i^t) - \log\left(1+\exp\left[\Phi_i^t(\bar{z}_i^t)\right]\right)\right]. \qquad [30]$$

We can then maximize this likelihood by adjusting the synaptic connections so as to achieve the target burst activity $\mathbf{B}^\star$. By differentiating with respect to the recurrent apical weights, we get:

$$\frac{\partial\mathcal{L}(\mathbf{B}^\star|\boldsymbol{w}^{s\to p})}{\partial w_{ij}^{s\to p}} = \left[\mathrm{B}_i^{\star,t+1} - p(\mathrm{B}_i^{t+1}=1)\right]\bar{z}_i^t e_j^t, \qquad [31]$$

where we have introduced the following two quantities:

$$p(\mathrm{B}_i^{t+1}=1) = \frac{\exp\left[\Phi_i^t(\bar{z}_i^t)\right]}{1+\exp\left[\Phi_i^t(\bar{z}_i^t)\right]} \quad \text{and} \quad e_j^t = \frac{\partial u_i^t}{\partial w_{ij}^{s\to p}}, \qquad [32]$$

respectively, a sigmoid probability for Φ_i^t and the spike response function (12). Given the basal window state $\bar{z}_i^t$, the target burst sequence is uniquely defined by the input projected to the apical distal compartment and can be written as $\mathrm{B}_i^{\star,t+1} = \bar{z}_i^t a_i^{\star,t+1}$. If we take the deterministic limit of the model ($\delta v \to 0$, where $p(\mathrm{B}_i^{t+1}=1) = \bar{z}_i^t a_i^{t+1}$) and then note that $\bar{z}_i^t\bar{z}_i^t = \bar{z}_i^t$, we can rewrite the previous expression in a cleaner form:

$$\frac{\partial\mathcal{L}(\mathbf{B}^\star|\boldsymbol{w}^{s\to p})}{\partial w_{ij}^{s\to p}} = \left[a_i^{\star,t+1} - a_i^{t+1}\right]\bar{z}_i^t e_j^t. \qquad [33]$$

This means that the spikes a_i^{t+1} in the proximal apical compartment should mimic the ones in the distal compartment, $a_i^{\star,t+1}$, when the somatic window $\bar{z}_i^t$ is open, in order to maximize the probability that the bursts generated by the network reproduce the target ones. The nondeterministic version of the learning rule (finite $\delta v = 0.1$), which is the one actually implemented in the experiments, reads:

$$\frac{\partial\mathcal{L}(\mathbf{B}^\star|\boldsymbol{w}^{s\to p})}{\partial w_{ij}^{s\to p}} = \left[a_i^{\star,t+1} - p\left(a_i^{t+1}=1|u_i^t\right)\right]\bar{z}_i^t e_j^t, \qquad [34]$$

where:

$$p\left(a_i^{t+1}=1|u_i^t\right) = \frac{\exp\left(\frac{u_i^t-v_{\mathrm{thr}}}{\delta v}\right)}{1+\exp\left(\frac{u_i^t-v_{\mathrm{thr}}}{\delta v}\right)}. \qquad [35]$$

We stress here how in the derivation we considered the basal-window state $\bar{z}_i^t$ as given. Consequently, the target burst sequence $\mathbf{B}^\star$ is uniquely defined by apical distal spikes $\boldsymbol{a}^\star$, which in turn depends on the input projected to the apical distal compartment. This sequence of spikes can be associated with a well-defined likelihood. Though we are aware of the feedback influence of the burst activity on the basal-window configuration (bursts induce basal spikes, see Eq. **25**), we chose to neglect such contribution as it would have severely increased the difficulty of the derivation. The convergence to the chosen target thus cannot be mathematically guaranteed. However, despite this limitation, it is important to note that the pattern of apical spikes $\boldsymbol{a}^\star$ does not change during learning: It is entirely determined by the original teaching signal $\boldsymbol{y}^\star$ and the variance σ_{targ} of its random projection to the network. As target bursts only occur after coincidence of an apical spike $a_i^\star$ and a basal spike z_i, the pattern $\mathbf{B}^\star$ is necessarily a subset of the fixed distal apical spikes $\boldsymbol{a}^\star$, and thus cannot diverge. In principle, it is still possible that the target pattern oscillates between slightly different subsets of $\boldsymbol{a}^\star$. In practice, we do not observe such behavior and provide numerical evidence that the target pattern of bursts converges to a well-defined pattern (see *SI Appendix* for details).

Discussion

It is more and more evident that dendrites are able to produce spikes (29) and perform complex and nonlinear computations (10). A famous example is the capability to initiate broad calcium action potentials ("Ca^{2+} spikes") near the apical tuft of pyramidal layer-5 neurons, that produce a long (up to 50 ms in vitro) plateau-type depolarization (11). The coincidence between this phenomenon and a somatic spike induces high-frequency somatic bursts during such depolarization.

In the present work, we model such a mechanism through the variable $\bar{\mathrm{B}}_i^t$ (Eq. **22**), that is, 1 for a 30 ms time window, after the coincidence between the apical (a_i^t) and the somatic (z_i^t) spikes. We show that this mechanism enables pyramidal neurons to naturally support target-based learning, that is easily applicable to, e.g., store-and-recall tasks. Moreover, it makes possible to use contextual signals to flexibly select the desired output from a repertoire of learned dynamics, acting as a hint or suggestion.

Also, we argue that this framework provides a natural solution to a general problem in learning: the plasticity–stability dilemma. A neural network requires to quickly capture statistical irregularities and learn new information, still retaining network stability to prevent forgetting previous memories. A first instantiation of this problem is Hebbian plasticity, that provides a positive feedback loop and leads to unstable runaway activity (30). To solve this, it has been suggested that homeostatic processes keep the network activity stable (31). In ref. 32, the authors show that the gating of plasticity in dendrites can improve network stability without compromising plasticity. This problem was also faced in a network of bursting neurons (33), where the authors define a homeostatic rule to regulate the bursting and the firing activity to a target value. We argue that our target-based approach intrinsically solves the plasticity–stability dilemma. Indeed, the learning instability is mitigated by an intrinsic separation of timescales between the proposed target (that slowly changes in time, increasing stability and preventing activity runaway), and the quick changes in synaptic weights (allowing a quick learning of proposed targets).

The neuronal architecture we propose allows to build hierarchically organized networks, which in turn are suited to orches-

trate HIL (20). It enables the decomposition of challenging long-horizon decision-making tasks into simpler subtasks, improving both learning speed and transfer learning, as skills learned by submodules can be reused for different tasks. In our work, a high-level network (the "manager") selects the correct policy for the task, suggesting it as a contextual signal to the low-level network (the "worker"), in charge of actually executing it. We also show how considering contextual information as an input for the apical compartment (instead of the basal one) is crucial for the correct decomposition and accomplishment of the task, in agreement with the biological interpretation of apical dendritic inputs as contextual signals from other cortical areas.

Though our hierarchical imitation learning approach requires devising handcrafted solutions for the different layers of the network, our message is that the architecture we propose can efficiently support the implementation of hierarchical policies to solve complex but decomposable tasks. In future works, we plan to replace behavioral cloning with more general learning schemes, such as feudal networks (34), where the high network (the manager) moves in a latent space, while the low network (the worker) translates its position into meaningful behavioral primitives.

To our knowledge, no other existing works propose a biologically plausible architecture to implement HIL. Furthermore, our model prepares the ground for further biological explorations. Tuning model parameters (e.g., the adaptation strength b) allows simulating the transition between different brain states (e.g., sleep and awake) (35–37). Possible future investigation topics include the replay of patterns of bursts during sleep (38), and the effect of sleep on tasks performances (35, 39–41).

Data, Materials, and Software Availability. Code for simulations and scripts for figures have been deposited in LTTB repository on GitHub (https://github.com/cristianocapone/LTTB) (42). All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. This work has been cofunded by the European Next Generation EU grants CUP I53C22001400006 (FAIR PE0000013 PNRR Project) and CUP B51E22000150006 (EBRAINS-Italy IR00011 PNRR Project) and by the European Union Horizon 2020 Research and Innovation program under the FET Flagship Human Brain Project (grant agreement SGA3 n. 945539).

Author affiliations: [a]Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Roma, Rome 00185, Italy; and [b]Scuola Internazionale Superiore di Studi Avanzati (SISSA), Visual Neuroscience Lab, Trieste 34136, Italy

Author contributions: C.C., C.L., and P.S.P. designed research; C.C., C.L., and P.M. performed research; C.C. formalized the theory and the model; and C.C., C.L., P.M., and P.S.P. wrote the paper.

1. W. Maass, Networks of spiking neurons: The third generation of neural network models. *Neural Netw.* **10**, 1659–1671 (1997).
2. G. Bellec *et al*., A solution to the learning dilemma for recurrent networks of spiking neurons. *Nat. Commun.* **11**, 1–15 (2020).
3. P. Muratore, C. Capone, P. S. Paolucci, Target spike patterns enable efficient and biologically plausible learning for complex temporal tasks. *PLoS One* **16**, e0247014 (2021).
4. C. Capone, P. Muratore, P. S. Paolucci, Error-based or target-based? A unified framework for learning in recurrent spiking networks *PLoS Comput. Biol.* **18**, e1010221 (2022).
5. Y. Cao, Y. Chen, D. Khosla, Spiking deep convolutional neural networks for energy-efficient object recognition. *Int. J. Comput. Vis.* **113**, 54–66 (2015).
6. K. Yamazaki, V.-K. Vo-Ho, D. Bulsara, N. Le, Spiking neural networks and their applications: A review. *Brain Sci.* **12**, 863 (2022).
7. W. Dengyu, X. Yi, X. Huang, A little energy goes a long way: Build an energy-efficient, accurate spiking neural network from convolutional neural network. *Front. Neurosci.* **16**, 759900 (2022).
8. W. Nicola, C. Clopath, Supervised learning in spiking neural networks with force training. *Nat. Commun.* **8**, 2208 (2017).
9. A. Payeur, J. Guerguiev, F. Zenke, B. A. Richards, R. Naud, Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nat. Neurosci.* **24**, 1010–1019 (2021).
10. P. Poirazi, A. Papoutsi, Illuminating dendritic function with computational models. *Nat. Rev. Neurosci.* **21**, 303–321 (2020).
11. M. Larkum, A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex. *Trends Neurosci.* **36**, 141–151 (2013).
12. R. Urbanczik, W. Senn, Learning by the dendritic prediction of somatic spiking. *Neuron* **81**, 521–528 (2014).
13. J. Guerguiev, T. P. Lillicrap, B. A. Richards, Towards deep learning with segregated dendrites. *Elife* **6**, e22901 (2017).
14. J. Sacramento, R. P. Costa, Y. Bengio, W. Senn, "Dendritic cortical microcircuits approximate the backpropagation algorithm" in *Advances in Neural Information Processing Systems 31*, S. Bengio *et al*., Eds. (Curran Associates Inc., 2018), pp. 8721–8732.
15. D.-H. Lee, S. Zhang, A. Fischer, Y. Bengio, "Difference target propagation" in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, A. Appice, P. Rodrigues, V. Santos Costa, C. Soares, J. Gama, A Jorge, Eds. (Springer, Cham, 2015), pp. 498–515.
16. B. DePasquale, C. J. Cueva, G. Kanaka Rajan, S. Escola, L. F., Abbott, full-FORCE: A target-based method for training recurrent networks. *PLoS One* **13**, e0191527 (2018).
17. N. Manchev, M. W. Spratling, Target propagation in recurrent neural networks. *J. Mach. Learn. Res.* **21**, 7–1 (2020).
18. A. Meulemans, F. Carzaniga, J. Suykens, J. Sacramento, B. F. Grewe, A theoretical framework for target propagation. *Adv. Neural. Inf. Process. Syst.* **33**, 20024–20036 (2020).
19. Allen brain atlas (2023). https://celltypes.brain-map.org/experiment/morphology/571735073.
20. H. Le *et al*., "Hierarchical imitation and reinforcement learning" in *International Conference on Machine Learning* (PMLR, 2018), pp. 2917–2926.
21. S. Pateria, B. Subagdja, A. Tan, C. Quek, Hierarchical reinforcement learning: A comprehensive survey. *ACM Comput. Surv. (CSUR)* **54**, 1–35 (2021).
22. L. J. Cauller, A. T. Kulics, The neural basis of the behaviorally relevant N1 component of the somatosensory-evoked potential in SI cortex of awake monkeys: Evidence that backward cortical projections signal conscious touch sensation. *Exp. Brain Res.* **84**, 607–619 (1991).
23. Minigrid gym: Door-key environment (2019). https://github.com/mit-acl/gym-minigrid/door--key-environment.
24. Minihack key-room environment (2021). https://minihack.readthedocs.io/en/latest/envs/navigation/keyroom.html.
25. B. Cramer *et al*., Surrogate gradients for analog neuromorphic computing. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2109194119 (2022).
26. J.-P. Pfister, T. Toyoizumi, D. Barber, W. Gerstner, Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Comput.* **18**, 1318–1348 (2006).
27. D. Jimenez Rezende, W. Gerstner, Stochastic variational learning in recurrent spiking networks. *Front. Comput. Neurosci.* **8**, 38 (2014).
28. B. Gardner, A. Grüning, Supervised learning in spiking neural networks for precise temporal encoding. *PLoS One* **11**, e0161335 (2016).
29. S. Gasparini, M. Migliore, J. C. Magee, On the initiation and propagation of dendritic spikes in CA1 pyramidal neurons. *J. Neurosci.* **24**, 11046–11056 (2004).
30. L. F. Abbott, S. B. Nelson, Synaptic plasticity: Taming the beast. *Nat. Neurosci.* **3**, 1178–1183 (2000).
31. G. G. Turrigiano, S. B. Nelson, Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.* **5**, 97–107 (2004).
32. K. A. Wilmes, C. Clopath, Dendrites help mitigate the plasticity-stability dilemma. *SSRN* (2022). 10.2139/ssrn.4121844.
33. F. Vercruysse, R. Naud, H. Sprekeler, Self-organization of a doubly asynchronous irregular network state for spikes and bursts. *PLoS Comput. Biol.* **17**, e1009478 (2021).
34. A. S. Vezhnevets *et al*., "Feudal networks for hierarchical reinforcement learning" in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup, Y. W. Teh, Eds. (PMLR, 2017), vol. 70, pp. 3540–3549.
35. Y. Wei, G. P. Krishnan, M. Komarov, M. Bazhenov, Differential roles of sleep spindles and sleep slow oscillations in memory consolidation. *PLoS Comput. Biol.* **14**, e1006322 (2018).
36. J. Goldman *et al*., *Brain-scale emergence of slow-wave synchrony and highly responsive asynchronous states based on biologically realistic population models simulated in the virtual brain*. BioRxiv (2020). https://www.biorxiv.org/content/10.1101/2020.12.28.424574v1 (Accessed 7 November 2023).
37. N. Tort-Colet, C. Capone, M. V. Sanchez-Vives, M. Mattia, Attractor competition enriches cortical dynamics during awakening from anesthesia. *Cell Rep.* **35**, 109270 (2021).
38. K. Kaefer, M. Nardin, K. Blahna, J. Csicsvari, Replay of behavioral sequences in the medial prefrontal cortex during rule switching. *Neuron* **106**, 154–165 (2020).
39. C. Capone, E. Pastorelli, B. Golosio, P. S. Paolucci, Sleep-like slow oscillations improve visual classification through synaptic homeostasis and memory association in a thalamo-cortical model. *Sci. Rep.* **9**, 1–11 (2019).
40. B. Golosio *et al*., Thalamo-cortical spiking model of incremental learning combining perception, context and NREM-sleep. *PLoS Comput. Biol.* **17**, e1009045 (2021).
41. C. Capone, P. S. Paolucci, Towards biologically plausible dreaming and planning. arXiv [Preprint] (2022). http://arxiv.org/abs/2205.10044 (Accessed 7 November 2023).
42. C. Capone, C. Lupo, P. Muratore, P. S. Paolucci, LTTB. Github. https://github.com/cristianocapone/LTTB (Accessed 7 November 2023).