Interneuron diversity is required for compartment-specific feedback inhibition

Joram Keijser[1,2] and Henning Sprekeler[1, 3, *]

[1]Modelling of Cognitive Processes, Institute of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, 10587 Berlin, Germany

[2]Charité – Universitätsmedizin Berlin, Einstein Center for Neurosciences Berlin, 10117, Berlin, Germany

[3]Bernstein Center for Computational Neuroscience Berlin, 10115 Berlin, Germany

[*]Correspondence addressed to h.sprekeler@tu-berlin.de

Abstract

Cortical inhibitory interneurons consist of many subtypes that have been associated with different functions. Here we use an optimization approach to show that two classes of interneurons are necessary to implement compartment-specific feedback inhibition to pyramidal cells. The two classes resemble PV-expressing and SST-expressing interneurons in their connectivity and short-term plasticity, suggesting a functional role for their diverse characteristics.

1 Main Text

Cortical inhibitory interneurons are very diverse [1]. The two most common interneuron classes—parvalbumin (PV) positive and somatostatin (SST) positive cells—differ prominently in their connectivity and synaptic dynamics: whereas PV basket cells typically receive short-term depressing input from excitatory pyramidal cells and in turn inhibit their soma and proximal dendrites, SST Martinotti cells receive short-term facilitating input and inhibit distal dendrites [2]. But what is the function of these differences between PV and SST interneurons?

One of inhibition's core functions is to prevent run-away excitation [3] by means of feedback inhibition that tracks excitatory inputs. This has led to the concept of excitation-inhibition (E/I) balance [4]. E/I balance is thought to shape cortical dynamics [4] and computations [5, 6] and can be established by means of inhibitory forms of plasticity [7]. Selective disruptions of E/I balance are thought to play a key role during learning [8].

Originally conceived as a balance on average [4], E/I balance turned out to be specific to sensory stimuli [9], in time [10, 11], across neurons [12] and to neural activation patterns [13]. Given the high specificity of E/I balance, we hypothesized that excitation and inhibition also balance across different neuronal compartments [14], and that this could be mediated at least in part by compartment-specific feedback inhibition. Different neuronal compartments often receive input from different sources [15] and display complex nonlinear dynamics [16, 17] that shape how these inputs are integrated [18]. We hypothesized that compartment-specific feedback inhibition requires a reflection of this intracellular complexity in the surrounding inhibitory circuitry. In particular, we hypothesized that it creates a need for different interneuron classes. If this were true, interneuron classes should emerge in a recurrent network model that is optimized for a compartment-specific E/I balance.

To test this idea, we simulated spiking networks comprising pyramidal cells (PCs) and interneurons (INs) (see Methods). The PCs were described by a two-compartment model consisting of a soma and an apical dendrite. The parameters of this model were previously fitted to capture dendrite-dependent bursting [19]. PCs received time-varying inputs in both the somatic and the dendritic compartment. INs were described by point neurons that receive excitatory inputs from the PCs and return feedback inhibition to the PCs. The strength of all synaptic connections in the network and the short-term plasticity of the PC $\rightarrow$ IN connections were optimized to balance excitation and inhibition across PC compartments, by means of gradient descent ([20], see Methods).

Before optimization, interneurons formed a single, homogeneous group (Fig. 1a, top). Most inhibited both somatic and dendritic compartments (Fig. 1b, top) and PC $\rightarrow$ IN connections showed non-specific synaptic dynamics (Fig. 1c, top). Moreover, excitation and inhibition were poorly correlated, particularly in the dendrite (Pearson correlation coefficients 0.55 (soma) & 0.04 (dendrite)), suggesting that the network did not generate compartment-specific feedback inhibition (Fig. 1d, top).

During optimization, the interneurons robustly split into two groups (Fig. 1a, bottom) with different connectivity (Fig. 1b, bottom) and short-term plasticity (Fig. 1c, bottom). One group received short-term depressing inputs from PCs and preferentially targeted their somatic compartment, akin to PV interneurons. The other group received short-term facilitating inputs from PCs and targeted their dendritic compartment, akin to SST interneurons. Excitation and inhibition were now positively correlated in both compartments (Pearson correlation coefficients 0.77 (soma) & 0.61 (dendrite); Fig. 1d, bottom).

To confirm the benefit of non-overlapping interneuron classes, we performed control simulations in which each interneuron was pre-assigned to target either the soma or the dendrite, while synaptic strengths and short-term plasticity were optimized. Consistent with a benefit of a specialization, the correlation of excitation and inhibition in the two compartments was as high as in fully self-organized networks (Fig. A.1). Optimized networks with pre-assigned interneuron classes also showed the same diversification in their short-term plasticity, resembling that of PV and SST neurons (Figs. A.1, A.2).

Because interneurons subtypes also differ in their connectivity to other interneurons [21, 22], we included IN $\rightarrow$ IN synapses in our optimization. After classifying INs as putative PV and SST neurons using a binary Gaussian mixture model, we found that the connections between the interneuron classes varied systematically

in strength. While PV↔PV connections, PV→SST connections and SST↔SST connections were similar in strength on average, SST→PV were consistently stronger (Fig. A.3a). To investigate which connections were necessary, we performed simulated knockout experiments in networks with pre-assigned interneuron classes, in which we removed individual connections types. We found that only PV→SST connections were necessary for a dendritic E/I balance (Fig. A.3b). This was confirmed by a mathematical analysis of a simplified network model, which suggests that disynaptic PC→PV→SST inhibition is necessary to prevent somatic inputs from generating dendritic inhibition (Section B.1). Although earlier work did not find PV→SST connectivity in the primary visual cortex of young mice [21], these connections are present in primary visual and somatosensory cortex of older animals [22].

For compartment-specific feedback inhibition, interneurons have to retrieve the somatic and dendritic input to PCs from the spiking activity of the PCs (Fig. 2a). By which mechanism is this decoding achieved? Recently, it was proposed that the electrophysiological properties of PCs support a multiplexed neural code that simultaneously represents somatic and dendritic inputs in temporal spike patterns ([23], Fig. 2b). In this code, somatic input increases the number of spikes (event rate, see Methods), whereas dendritic input increases the probability that a spike is converted to a burst (burst probability). Providing soma- or dendrite-specific inhibition then amounts to decoding the event rate or burst probability, respectively. Such a decoding can be achieved in circuits with short-term plasticity and feedforward inhibition [23], and we expected that our network arrived at a similar decoding scheme.

We tested this hypothesis by injecting current pulses to PC somata and dendrites (see Methods). Stronger dendritic input increased the burst probability, which increased the SST rate, which increased dendritic inhibition (Fig. 2c-e, top). Analogously, stronger somatic input increased the event rate, which increased the PV rate, which increased somatic inhibition (Fig. 2c-e, bottom). Importantly, inhibition was specific to each compartment: Because PV neurons were selectively activated by PC events, somatic inhibition was largely unaffected by dendritic excitation. Similarly, SST neurons were selectively activated by PC bursts, such that dendritic inhibition was largely unaffected by somatic excitation (Fig. A.4). In the model, interneurons therefore provide compartment-specific inhibition by demultiplexing the neural code used by the PCs. Note that the balance is less tight in time in the dendrites than in the somata (cf. Fig. 2e top and bottom), a consequence of the delay between burst onset and SST activation [24].

So far we assumed that PC somata and dendrites receive uncorrelated input. Recent work, however, suggests that somatic and dendritic activity are correlated [25, 26], potentially reducing the need for

compartment-specific inhibition. We therefore tested how correlated inputs affect interneuron specialization by optimizing separate networks for different input correlations. We found that increasing correlation between somatic and dendritic inputs gradually reduced the separation between the interneuron classes (Fig 3a,b). For high input correlation, optimized networks contained a continuum in their connectivity and short-term plasticity (Fig. 3a,b). However, the presence of short-term plasticity was necessary for a dendritic E/I balance for a range of input correlations (Fig. 3c). Note that although distinct interneuron populations were not necessary for the case of high input correlation, the presence of IN classes was not harmful for E/I balance. A pre-assignment of the interneurons into classes maintained the E/I correlation in both compartments and for any correlation level (Fig A.2a). Finally, we found that interneuron specialization degraded with increasing baseline activity of the INs (Fig A.5), because high firing rates allow non-specialized inhibition to cancel out (see mathematical analysis in Section B.2). However, a pre-assignment of interneurons into classes again maintained the E/I correlation for different baseline activity levels (Fig A.2c).

The model predicts that, first, PV and SST rates should correlate with somatic and dendritic activity, respectively (Fig. 2). Second, inhibiting SST neurons [27] or manipulating PC $\rightarrow$ SST facilitation [28] should increase PC bursting. On a higher level, the model suggests a relation between the biophysical properties of excitatory neurons and the surrounding interneuron circuity. This is consistent, e.g., with the finding that the prevalence of pyramidal cells and dendrite-targeting Martinotti cells seems to be correlated across brain regions [29]. Combined with more complex biophysical models, the suggested optimization approach could hence provide an inroad to understanding further properties of interneuron classes and their circuitry.

Acknowledgements

J.K. was supported by a PhD scholarship from the Einstein Center for Neurosciences Berlin. We thank Loreen Hertäg, Robert Tjarko Lange, Laura Bella Naumann, and Filip Vercruysse for comments on earlier drafts.

Author contribution

J.K and H.S. designed the study. J.K. performed the experiments and analyzed the data. J.K. and H.S. wrote the manuscript.

Declaration of interests

The authors declare no competing interests.

Code availability

Code and trained models will be made publicly available upon publication.

2 References

[1] Robin Tremblay, Soohyun Lee, and Bernardo Rudy. Gabaergic interneurons in the neocortex: from cellular properties to circuits. *Neuron*, 91(2):260–292, 2016.

[2] Alex Reyes, Rafael Lujan, Andrej Rozov, Nail Burnashev, Peter Somogyi, and Bert Sakmann. Target-cell-specific facilitation and depression in neocortical circuits. *Nature neuroscience*, 1(4):279–285, 1998.

[3] Jeffry S Isaacson and Massimo Scanziani. How inhibition shapes cortical activity. *Neuron*, 72(2):231–243, 2011.

[4] Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

[5] Guillaume Hennequin, Tim P Vogels, and Wulfram Gerstner. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, 82(6):1394–1406, 2014.

[6] Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417, 2015.

[7] Tim P Vogels, Henning Sprekeler, Friedemann Zenke, Claudia Clopath, and Wulfram Gerstner. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334(6062):1569–1573, 2011.

[8] Johannes J Letzkus, Steffen BE Wolff, and Andreas Lüthi. Disinhibition, a circuit mechanism for associative learning and memory. *Neuron*, 88(2):264–276, 2015.

[9] Michael Wehr and Anthony M Zador. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426(6965):442–446, 2003.

[10] Michael Okun and Ilan Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience*, 11(5):535–537, 2008.

[11] Alfonso Renart, Jaime De La Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D Harris. The asynchronous state in cortical circuits. *science*, 327(5965):587–590, 2010.

[12] Mingshan Xue, Bassam V Atallah, and Massimo Scanziani. Equalizing excitation–inhibition ratios across visual cortical neurons. *Nature*, 511(7511):596–600, 2014.

[13] Aanchal Bhatia, Sahil Moza, and Upinder Singh Bhalla. Precise excitation-inhibition balance controls gain and timing in the hippocampus. *Elife*, 8:e43415, 2019.

[14] Daniel Maxim Iascone, Yujie Li, Uygar Sümbül, Michael Doron, Hanbo Chen, Valentine Andreu, Finola Goudy, Heike Blockus, Larry F Abbott, Idan Segev, et al. Whole-neuron synaptic mapping reveals spatially precise excitatory/inhibitory balance limiting dendritic and somatic spiking. *Neuron*, 2020.

[15] Leopoldo Petreanu, Tianyi Mao, Scott M Sternson, and Karel Svoboda. The subcellular organization of neocortical excitatory connections. *Nature*, 457(7233):1142–1145, 2009.

[16] Michael London and Michael Häusser. Dendritic computation. *Annu. Rev. Neurosci.*, 28:503–532, 2005.

[17] Panayiota Poirazi and Athanasia Papoutsi. Illuminating dendritic function with computational models. *Nature Reviews Neuroscience*, pages 1–19, 2020.

[18] Matthew Larkum. A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in neurosciences*, 36(3):141–151, 2013.

[19] Richard Naud, Brice Bathellier, and Wulfram Gerstner. Spike-timing prediction in cortical neurons with active dendrites. *Frontiers in computational neuroscience*, 8:90, 2014.

[20] Friedemann Zenke and Surya Ganguli. Superspike: Supervised learning in multilayer spiking neural networks. *Neural computation*, 30(6):1514–1541, 2018.

[21] Carsten K Pfeffer, Mingshan Xue, Miao He, Z Josh Huang, and Massimo Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience*, 16(8):1068–1076, 2013.

[22] Mahesh M Karnani, Jesse Jackson, Inbal Ayzenshtat, Jason Tucciarone, Kasra Manoocheri, William G Snider, and Rafael Yuste. Cooperative subnetworks of molecularly similar interneurons in mouse neocortex. *Neuron*, 90(1):86–100, 2016.

[23] Richard Naud and Henning Sprekeler. Sparse bursts optimize information transmission in a multiplexed neural code. *Proceedings of the National Academy of Sciences*, 115(27):E6329–E6338, 2018.

[24] Frederic Pouille and Massimo Scanziani. Routing of spike series by dynamic circuits in the hippocampus. *Nature*, 429(6993):717–723, 2004.

[25] Lou Beaulieu-Laroche, Enrique HS Toloza, Norma J Brown, and Mark T Harnett. Widespread and highly correlated somato-dendritic activity in cortical layer 5 neurons. *Neuron*, 103(2):235–241, 2019.

[26] Valerio Francioni, Zahid Padamsey, and Nathalie L Rochefort. High and asymmetric somato-dendritic coupling of v1 layer 5 neurons independent of visual stimulation and locomotion. *Elife*, 8:e49145, 2019.

[27] Sébastien Royer, Boris V Zemelman, Attila Losonczy, Jinhyun Kim, Frances Chance, Jeffrey C Magee, and György Buzsáki. Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nature neuroscience*, 15(5):769–775, 2012.

[28] Emily L Sylwestrak and Anirvan Ghosh. Elfn1 regulates target-specific release probability at ca1-interneuron synapses. *Science*, 338(6106):536–540, 2012.

[29] Federico Scala, Dmitry Kobak, Shen Shan, Yves Bernaerts, Sophie Laturnus, Cathryn Rene Cadwell, Leonard Hartmanis, Emmanouil Froudarakis, Jesus Ramon Castro, Zheng Huan Tan, et al. Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas. *Nature communications*, 10(1):1–12, 2019.

[30] Misha Tsodyks, Klaus Pawelzik, and Henry Markram. Neural networks with dynamic synapses. *Neural computation*, 10(4):821–835, 1998.

[31] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[32] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[33] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318, 2013.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[35] Matthew E Larkum, J Julius Zhu, and Bert Sakmann. A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398(6725):338–341, 1999.

3 Figures

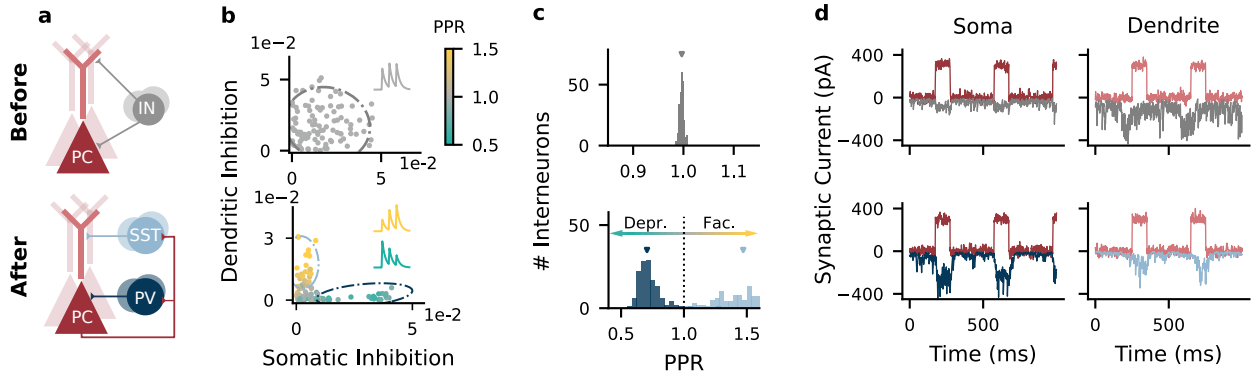

Figure 1: Interneuron diversity emerges in networks optimized for compartment-specific inhibition **(a)** Network structure before (top) and after optimization (bottom). PC, pyramidal cell; IN, interneuron; PV, parvalbumin-positive IN; SST, somatostatin-positive IN. Recurrent inhibitory connections among INs omitted for clarity. **(b)** Strength of somatic and dendritic inhibition from individual INs. Dashed lines: 95% density of a Gaussian distribution (top) and mixture of two Gaussian distributions (bottom) fitted to the connectivity and Paired Pulse Ratio (PPR) data of 5 networks (marginalized over PPR). **(c)** PPR distribution (data from 5 networks). Mean PPR before optimization: 1.00; after optimization: 0.73 (PV cluster, $n = 133$) and 1.45 (SST cluster, $n = 113$). **(d)** Excitatory (red) and inhibitory (top: gray, bottom: blue) currents onto PC compartments (average across $N_E = 400$ PCs). Correlation between excitation and inhibition before optimization: 0.55 (soma) and 0.04 (dendrites). After optimization: 0.77 (soma) and 0.61 (dendrites).

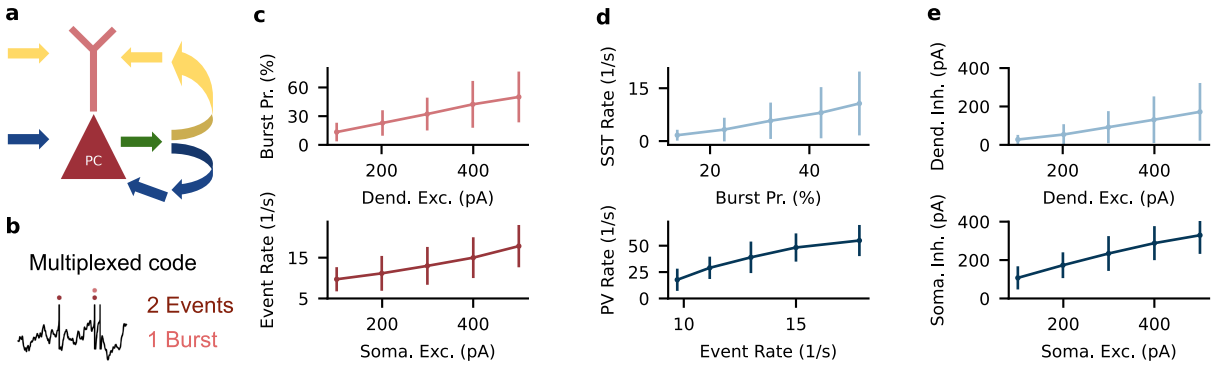

Figure 2: The interneuron circuit decodes somatic and dendritic inputs to PCs. **(a)** PC somata and dendrites receive uncorrelated input streams (yellow and blue) that, from PC output spikes (green), have to be separated into compartment-specific inhibition (yellow and blue). **(b)** PCs use a multiplexed neural code. Somatic input leads to events (singlets or bursts). Dendritic input converts singlets into bursts. **(c)** Top: Excitatory input to PC dendrites increases burst probability. Bottom: Excitatory input to PC somata increases event rate. Error bars indicate sd over 10 stimulus repetitions. **(d)** Top: SST rate increases with bursts probability. Bottom: PV rate increases with PC events. **(e)** Dendritic and somatic inhibition in PCs increase with dendritic and somatic excitation, respectively.

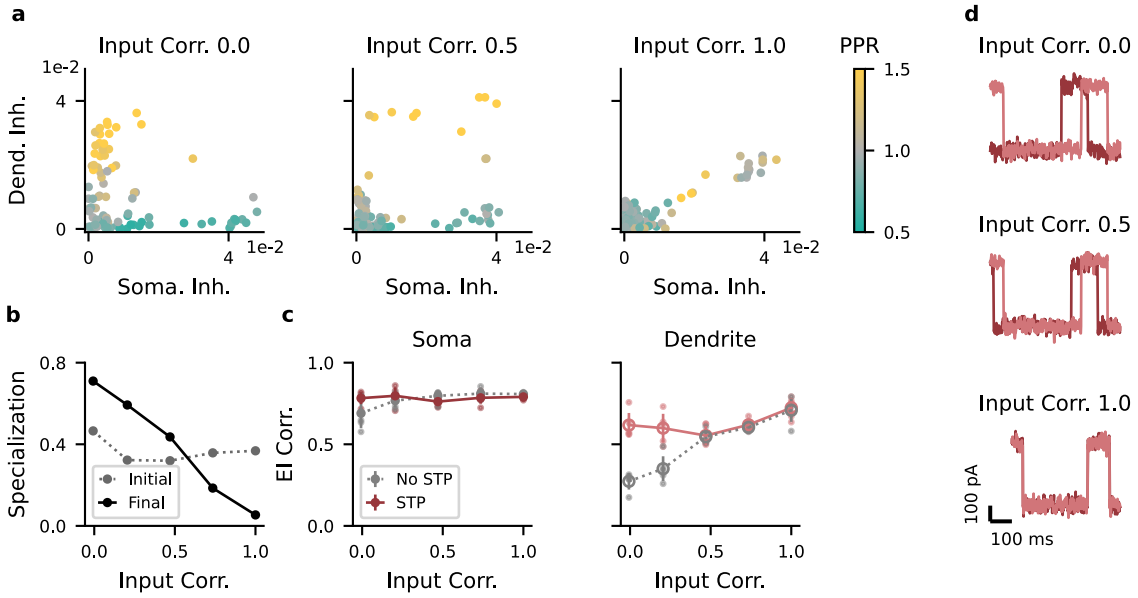

Figure 3: Correlations between dendritic and somatic input reduce interneuron specialization. **(a)** Strength of somatic vs. dendritic inhibition from all INs. Left, middle, right: input correlation coefficient 0 (low), 0.5 (medium), and 1 (high), respectively. **(b)** Specialization of IN→E weights. If each IN targets either soma or dendrites, the specialization is 1 (see Methods). Gray: specialization of initial random network; black: specialization after optimization. **(c)** Left: In the soma, excitation and inhibition are balanced across a broad range of input correlations, with or without short-term plasticity (STP). Right: In the dendrites, excitation and inhibition are balanced only with STP. Open circles, mean over 5 batches of 8 stimuli with random amplitudes (see Methods). Small filled circles, individual batches. **(d)** Examples for synaptic traces corresponding to correlation levels in **a**. Dark red, somatic current; light red, dendritic input.

4 Methods

4.1 Network Model

We simulated a fully connected spiking network model consisting of N_E pyramidal cells (PCs) and N_I interneurons (INs), as in earlier work [23]. PCs are described by a two-compartment model [19]. The membrane potential v^s in the somatic compartment is modeled as a leaky integrate-and-fire unit with spike-triggered adaptation:

$$\frac{dv^s}{dt} = -\frac{v^s - E_L}{\tau_s} + \frac{g_s f(v^d) + w^s + I^s}{C_s} \tag{1}$$

$$\frac{dw^s}{dt} = -\frac{w^s}{\tau_{s,w}} + b_s S(t)\,. \tag{2}$$

Here, E_L denotes the resting potential, τ_s the membrane time constant and C_s the capacitance of the soma. I^s is the external input, and w^s the adaptation variable, which follows leaky dynamics with time constant $\tau_{s,w}$, driven by the spike train S emitted by the soma. b_s controls the strength of the spike-triggered adaptation. v^d is the dendritic membrane potential, the conductance g_s controls how strongly the dendrite drives the soma, and f the nonlinear activation of the dendrite:

$$f(v) = 1/(1 + \exp(-(v - E_d)/D_d))\,. \tag{3}$$

The half-point E_d and slope D of the transfer function f control the excitability of the dendrite. When the membrane potential reaches the spiking threshold ϑ, it is reset to the resting potential and the PC emits a spike. Every spike is followed by an absolute refractory period of τ_r.

The dynamics of the dendritic compartment are given by:

$$\frac{dv^d}{dt} = -\frac{v^d - E_L}{\tau_d} + \frac{g_d f(v^d) + c_d K(t - \hat{t}) + w^d + I^d}{C_d} \tag{4}$$

$$\frac{dw^d}{dt} = -\frac{w^d}{\tau_{d,w}} + \frac{a_d(v^d - E_L)}{\tau_{d,w}}. \tag{5}$$

In addition to leaky membrane potential dynamics with time constant τ_d, the dendrite shows a voltage-dependent nonlinear activation f, the strength of which is controlled by g_d. This nonlinearity allows the generation of dendritic plateau potentials ("calcium spikes"). Somatic spikes trigger backpropagating action potentials in the dendrite, modeled in the form of a boxcar kernel K, which starts 1ms after the spike and lasts 2ms. The amplitude of the backpropagating action potential is controlled by the parameter c_d. The

dendrite is subject to a voltage-activated adaptation current w^d, which limits the duration of the plateau potential. This adaptation follows leaky dynamics with time constant $\tau_{d,w}$. The strength of the adaptation is given by the parameter a_d. Note that the model excludes sub-threshold coupling from the soma to the dendrite.

The interneurons are modeled as leaky integrate-and-fire neurons:

$$\frac{dv^i}{dt} = -\frac{v^i - E_L}{\tau_i} + \frac{I^i}{C_i}, \tag{6}$$

with time constant τ_i. Spike threshold, resting and reset potential, and refractory period are the same as for the PCs.

All neurons receive an external background current to ensure uncorrelated activity, which follows Ornstein-Uhlenbeck dynamics

$$\frac{dI^{x,bg}}{dt} = -\frac{I^{x,bg} - \mu_x}{\tau_{bg}} + \sigma_x \varepsilon. \tag{7}$$

Here, $x \in \{s, d, i\}$ refers to the soma, dendrite, or interneuron, respectively, and ε is standard Gaussian white noise with zero mean and correlation $\langle \varepsilon(t)\varepsilon(t') \rangle = \delta(t - t')$.

In addition, the somatic and dendritic compartments received step currents mimicking external signals (see Section 4.2), as well as recurrent inhibitory inputs. The recurrent input to compartment $x \in \{s, d\}$ of the ith principal cell was given by

$$I_i^{x,inh}(t) = -\sum_{j=1}^{N_I} |W_{ij}^{I \to x}| \, s^j(t). \tag{8}$$

where s^j is the synaptic trace that is increased at each presynaptic spike and decays with time constant τ_{syn} otherwise:

$$\frac{ds}{dt} = -\frac{s}{\tau_{syn}} + S.$$

The compartment-specific inhibitory weight matrices $W^{I \to x}, x \in \{s, d\}$ were optimized; the absolute value in Eq. 8 ensured positive weights.

The recurrent input to the ith interneuron was given by:

$$I_i^{rec} = \sum_{j=1}^{N_E} |W_{ij}^{E\to I}|\ \mu_{ij}(t)\ s^j(t) - \sum_{k=1} |W_{ik}^{I\to I}| s^k(t). \tag{9}$$

The function $\mu_{ij}(t)$ implements short-term plasticity according to the Tsodyks-Markram model [30]. $\mu(t)$ is the product of a utilization variable u and a recovery variable R that obey the dynamics

$$\frac{du}{dt} = -\frac{u-U}{\tau_u} + (1-u)\cdot F \cdot S, \tag{10}$$

$$\frac{dR}{dt} = -\frac{R-1}{\tau_R} - u \cdot R \cdot S. \tag{11}$$

U is the initial release probability, which is optimized by gradient descent. F is the facilitation fraction, and τ_R, τ_u are the time constants of facilitation and depression, respectively. All parameter values are listed in Table 1.

Finally, the network parameters were scaled so that the membrane voltages ranged between $E_L = 0$ and $\vartheta = 1$. The scaling allowed weights of order $1/\sqrt{N}$, mitigating vanishing or exploding gradients during optimization. All optimization parameters are listed in Table 2.

4.2 Optimization

We used gradient descent to find weights W and initial release probabilities U that minimize the difference between excitation and inhibition in both compartments:

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{i=1}^{N_E} \left(E^s(t) + I_i^s(t)\right)^2 + \left(E^d(t) + I_i^d(t)\right)^2. \tag{12}$$

E_i^x and I_i^x are the total excitatory and inhibitory input to compartment $x \in \{s, d\}$ of PC i. To speed up the optimization process, all output synapses from a given neuron to a given compartment type had the same strength, i.e., the optimization of the output synapses is performed for $N_I \times 2$ parameters. For the input synapses onto the INs, weight and initial release probability were optimized independent for all $N_E \times N_I$ synapses.

To achieve small interneuron rates necessary for interneuron specialization (Fig. A.5), we subtracted the

mean background input from E_i^x:

$$I_i^x(t) = E_i^x(t) - \mu_x \tag{13}$$

To propagate gradients through the spiking non-linearity, we replaced its derivative with the derivative of a smooth approximation [20]

$$\sigma(v) = \frac{1}{(1 + \beta|v - \vartheta|)^2}. \tag{14}$$

We used the machine learning framework PyTorch [31] to simulate the differential equations (forward Euler with step size 1 ms), compute the gradients of the objective $\mathcal{L}$ using automatic differentiation, and update the network parameters using Adam [32]. The optimized parameters were initialized according to the distributions listed in Table 2. We simulated the network response to batches of 8 trials of 600 ms, consisting of 100 ms pulses given at 2.5 Hz. The pulse amplitudes were drawn uniformly and independently for soma and dendrites from the set $\{100, 200, 300, 400\}$. Training converged within 200 batches (parameter updates). Before each parameter update, the gradient values were clipped between -1 and 1 to mitigate exploding gradients [33]. After each update, the initial release probability was clipped between 0 and 1 to avoid unphysiological values.

Symbol	Value	Unit	Description
N_E	400	-	Number of exc. neurons
N_I	100	-	Number of inh. neurons
E_L	-70	mV	reversal and reset potential
ϑ	-50	mV	spiking threshold
$\tau_{s/d/i}$	16 / 7 / 10	ms	time const. soma/ dend./inh. membrane
τ_r	3	ms	refractory time soma and inh.
$g_{s/d}$	1300 / 1200	pA	Coupling from dend to soma
$C_{s/d/i}$	370 /170/100	pF	Conductance of soma/dend./inh.
$\tau_{s/d,w}$	100 / 30	ms	Time const. adaptation soma/dend.
b_s	-200	pA	Spike-triggered adaptation (soma)
a_d	-13	nS	Voltage-driven adaptation (dend)
c_d	2600	pA	Coupling soma to dend.
E_d	-38	mV	position dend. nonlinearity
D_d	6	mV	steepness of dend. nonlinearity
$\mu_{s/d/i}$	400 / -300 /-100	pA	mean background input soma/dend./inh.
$\sigma_{s/d/i}$	450 / 450 / 400	pA	sd background input
τ_{bg}	2	ms	time const. background input
τ_{syn}	5	ms	time const. synapses
τ_u	100	ms	time const. facilitation
τ_R	100	ms	time const. depression
F	0.1	-	facilitation jump

Table 1: Parameter values for network simulation.

Symbol	Value / Init. Distribution	Dimensions	Description
U	$\mathcal{U}(0.1, .25)$	$N_E \times N_I$	Initial release prob.
$W^{E\to I}$	$\mathcal{N}(0, 1/N_E)$	$N_E \times N_I$	Exc. to Inh. weight
$W^{I\to I}$	$\mathcal{N}(0, 1/N_I)$	$N_I \times N_I$	Inh. to Inh. weight
$W^{I\to D}$	$\mathcal{N}(0, 0.2/N_I)$	$N_I \times 1$	Inh. to Exc. Dend. weight
$W^{I\to S}$	$\mathcal{N}(0, 0.2/N_I)$	$N_I \times 1$	Inh. to Exc. Soma weight
-	1e-3	-	learning rate for weights
-	4e-3	-	learning rate for U
β	10	-	Slope spiking derivative
-	1.0	-	Gradient (absolute value) clipping

Table 2: Optimization parameters. $\mathcal{U}(a, b)$, uniform distribution on the interval $(a, b]$; $\mathcal{N}(0, \sigma^2)$, normal distribution with mean 0 and variance σ^2.

4.3 Methods for Figures

4.3.1 Figure 1

We measured the short-term plasticity of PC $\to$ IN synapses by simulating their response to two EPSPs given 10 ms apart, a typical interspike interval within a burst. The PPR was computed as the ratio of the two EPSP amplitudes, such that a PPR > 1 indicates short-term facilitation and a PPR < 1 indicates short-term depression. The PPR of a single IN was defined as the mean PPR of all its excitatory afferents. Clustering of interneurons was done by fitting a single Gaussian (before optimization) or a mixture of two Gaussians (after optimization) to the three-dimensional distribution of inhibitory weights to the PC soma, to PC dendrites, and the PC$\to$IN Paired Pulse Ratio (PPR). Both models were fitted using Scikit-learn [34] on pooled data from five networks, trained from different random initializations. The density models where fitted on 246 interneurons that were active (firing rate higher than 1 spk/s) and had a medium to strong projection to either soma or dendrites (weight bigger than 0.01). The dashed lines in Fig. 1b illustrate the two-dimensional marginal distributions of the somatic and dendritic inhibition. All PCs received the same time-varying input currents, consisting of 100 ms pulses of 300 pA, given at a rate of 2.5 Hz. Correlations between compartment-specific excitation and inhibition were computed between the the currents to the PC compartments, averaged across all PCs in the network.

4.3.2 Figure 2

The definitions of burst rate, burst probability and event rate were taken from Naud & Sprekeler [23]: A burst was defined as multiple spikes occurring within 16 ms. The time of the first spike was taken as the time of the burst. An event was defined as a burst or a single spike. The instantaneous burst rate and event rate were computed by counting the number of bursts and events, respectively, in bins of 1ms and among

the population of PCs, and smoothing the result with a Gaussian filter (width: 2ms). The burst probability was defined as

$$\text{Burst Probability} = \frac{\text{Burst Rate}}{\text{Event Rate}} \times 100\%. \tag{15}$$

We injected current pulses of 100 ms duration to either soma or dendrite while injecting a constant current to the other compartment. Currents where varied in amplitude between 100 and 400 pA; the constant current was 0 pA. The figure shows the mean and standard deviation of the total network activity during 10 current pulses.

4.3.3 Figure 3

We varied the correlation between the inputs to soma and dendrites by generating repeating current pulses with different temporal offsets and optimized a network for each offset. The interneuron specialization was defined as

$$\text{specialization} = 1 - \frac{x^T y}{\|x\|\|y\|}\,, \tag{16}$$

where x and y are N_I-dimensional vectors containing the inhibitory weights onto soma and dendrites and $\|\cdot\|$ the L_2 norm. If each neuron inhibits either somata or dendrites, but not both, the specialization will be 1. If the weights are perfectly aligned (i.e., interneurons with a strong dendritic projection also have a strong somatic projection), the specialization will be 0. Here and in all figures, the EI correlation was computed as the correlation between the time series of the compartment-specific excitation and inhibition, after averaging across all PCs. Shown is the mean over 5 batches of 600 ms, where each batch consisted of 8 trials with amplitudes from $\{100, 200, 300, 400\}$ pA, sampled independently for soma and dendrites.

4.3.4 Figure A.1

Before optimization, we assigned interneurons to inhibit either PC somata or dendrites by fixing their weights onto the other compartment to zero. Half of the interneurons was assigned to inhibit the soma, the other half was assigned to inhibit the dendrites. Otherwise, weights and initial release probabilities were optimized as before.

4.3.5 Figure A.2

As for Fig. A.1, we assigned interneurons to inhibit either PC somata or dendrites. Here, we trained networks for different correlations between compartment-specific external inputs (cf. Fig. 3), and baseline activity levels (cf. Fig. A.5). We used the 10th percentile as a robust measure of minimum PV rate. The mean and sd PPR of the PV and SST populations computed over all INs that were active (rate larger than 1 spk/s) and provided a medium to strong inhibition to one PC compartment (weight bigger than 0.01).

4.3.6 Figure A.3

Figure A.3a shows the connectivity strength over five networks. We first used the Gaussian mixture models to assign INs to PV or SST clusters, and then computed the mean connectivity between and within clusters for each network. For A.3b, we trained networks with predefined interneuron populations to control the interneuron connectivity. Connections between populations were knocked out by fixing them to zero during and after optimization. EI correlations are computed for 5 batches of 600 ms, where each batch consisted of 8 trials with amplitudes from $\{100, 200, 300, 400\}$ pA, sampled independently for soma and dendrites.

4.3.7 Figure A.4

As in Fig. 2, we injected current pulses of 100ms duration to either soma or dendrite. Here, we injected a simultaneous pulses to the other compartment of amplitude $0, 200$ or 400 pA.

4.3.8 Figure A.5

The minimum rate of PV neurons was controlled indirectly, by varying the baseline inhibitory target current to the soma—A larger baseline requires a higher minimum PV rate. We varied the minimum inhibitory current by subtracting only a fraction α of the baseline excitatory current:

$$I^x(t) = E^x(t) - \alpha \cdot \mu_x, \tag{17}$$

cf. Eq. (13). In the simulations, we varied α between 1 and 0.8, leading to a minimum PV rate between 1 spk/s, and 9 spk/s.

A Supplementary Figures

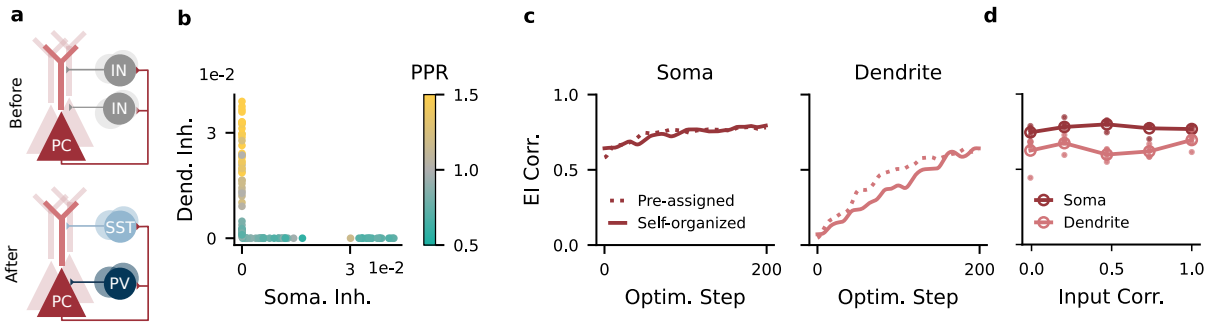

Figure A.1: Non-overlapping interneuron populations achieve compartment-specific inhibition. **(a)** Before optimization, interneurons are assigned to inhibit a single compartment. The optimization determines the synaptic strengths and the short-term plasticity. **(b)** Compartment-specific inhibition from active INs after optimization. Mean PPR: 0.78 (PV), 1.21 (SST). **(c)** Correlation between excitation and inhibition over the course of the optimization. Solid line: INs were not assigned to a single compartment (Self-organized). Dashed line: INs were assigned to a single compartment (Pre-assigned). Data is smoothed with a Gaussian kernel (width: 5). **(d)** EI correlation of pre-assigned networks for different correlation levels between compartment-specific external inputs.

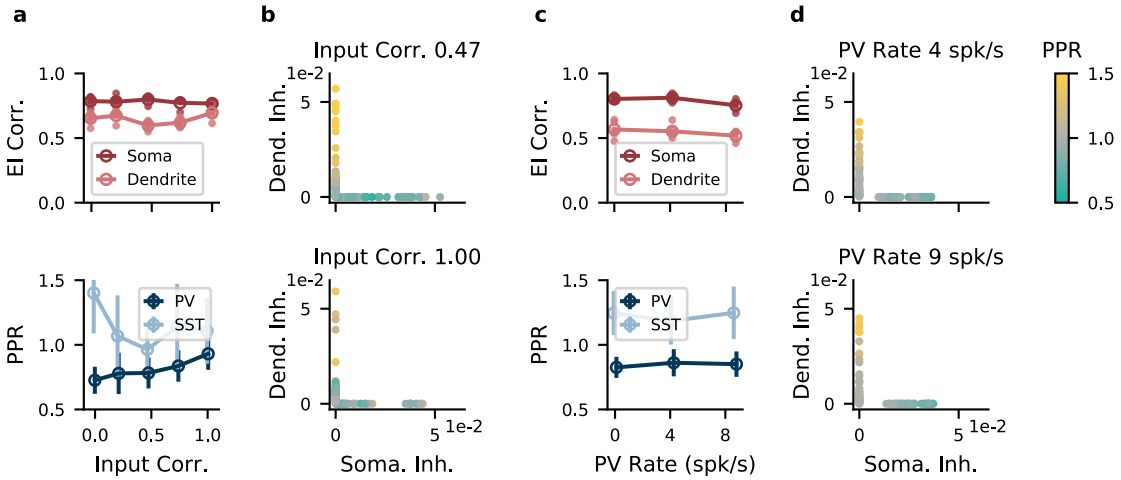

Figure A.2: Non-overlapping interneuron populations achieve compartment-specific inhibition for a range of input statistics. **(a)** Top, performance as measured by compartment-specific correlation between excitation and inhibition of networks trained on different correlations between compartment-specific excitatory inputs. Open circles, mean over 5 batches of 8 stimuli with random amplitudes (see Methods). Small filled circles, individual batches. Here and in the other panels, the interneurons were assigned to inhibit only the soma or only the dendrites. Bottom, interneuron specialization as measured by Paired Pulse Ratio (PPR) decreases with input correlations. Error bars denote sd over IN populations. **(b)** Strength of somatic and dendritic inhibition from individual INs. Top, medium input correlation (0.47); bottom, high input correlation (1.00). Color indicates PPR. **c)** Top, as **a** but as function of minimum PV rate. Bottom, interneuron specialization as measured by Paired Pulse Ratio (PPR) is not influenced by minimum PV rate. **(d)** Strength of somatic and dendritic inhibition from individual INs. Top, medium PV rate (4 spk/s); bottom, high PV rate (9 spk/s).

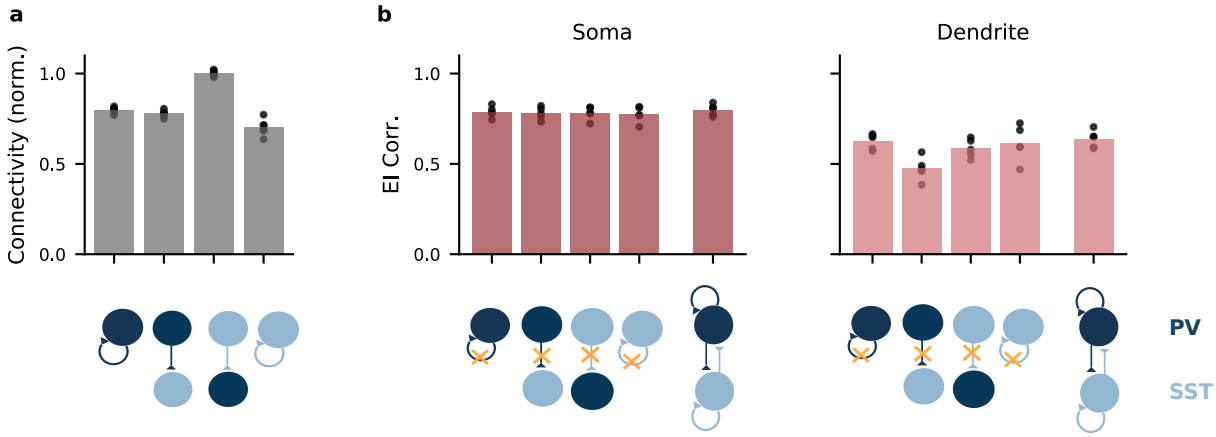

Figure A.3: Recurrent inhibitory connectivity after learning. **(a)** Connectivity between IN populations. From left to right: PV↔PV, PV→SST, SST→PV, SST↔SST. Bars indicate mean over all networks, dots indicate individual networks. **(b)** Performance as measured by the correlation between excitation and inhibition to PC soma (left) and dendrites (right) of networks optimized lacking specific connections. Data at the very right: EI correlation in network with unconstrained connectivity. Only loss of PV → SST connectivity has a clear effect on dendritic EI correlations. Open circles, mean over 5 batches of 8 stimuli with random amplitudes. Small filled circles, individual batches.

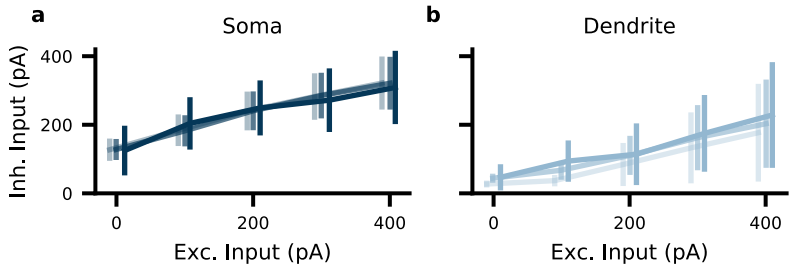

Figure A.4: Inhibition to one PC compartment is largely independent of excitation to the other compartment. **(a)** Somatic inhibition increases with somatic excitation, but is invariant to dendritic excitation. Shading indicates strength of somatic input; bright, medium, dark: 0, 200, and 400 pA, respectively. Positions on x-axis are shifted by 10 pA for visual clarity, error bars indicate sd during 10 stimulus repetitions. **(b)** Dendritic inhibition increases with dendritic excitation, but is only weakly modulated by somatic excitation. Shading indicates strength of dendritic input (0, 200, and 400 pA).

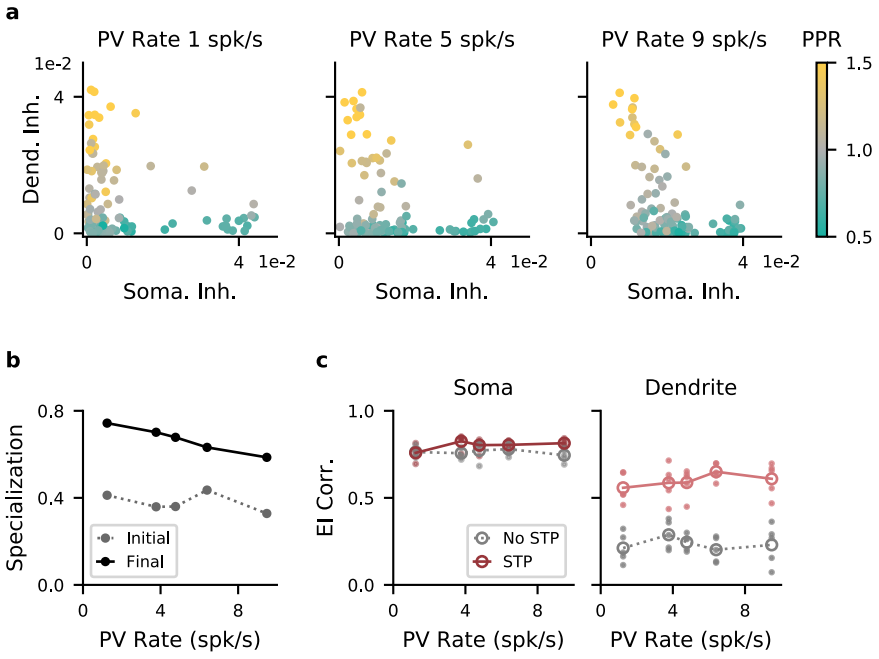

Figure A.5: Higher baseline PV rates decrease the need for interneuron specialization. **(a)** Strength of somatic and dendritic inhibition from individual INs. Left, middle, right: network optimized with a baseline PV rate of 1 (low), 5 (medium), and 9 spk/s (high), respectively. **(b)** Specialization of IN→E weights. If each IN targets either soma or dendrites, the specialization is 1 (see Methods). Gray: specialization of initial, random network; black: specialization after optimization. **(c)** Left, correlation between excitation and inhibition as function of minimum PV rate. Red: networks with optimized short-term plasticity. Gray: Networks without short-term plasticity. Open circles, mean over 5 batches of 8 stimuli with random amplitudes. Small filled circles, individual batches.

B Mathematical Analysis of a Simplified Network Model

We performed a mathematical analysis of a simplified network to better understand the following results of our spiking network simulations:

1. A compartment-specific balance requires PV → SST inhibition, but no other IN → IN connectivity (Fig. A.3).

2. Higher interneuron rates require less IN specialization, i.e., individual interneurons often inhibit both PC compartments (Fig A.5).

The simplified model consists of a population of principal cells (PC) and two populations of interneurons that we will refer to as parvalbumin (PV)-positive and as somatostatin (SST)-positive cells. The population activity of the PCs is represented by somatic activity e and dendritic activity b. The interneuron activities are represented by firing rates p and s. The four activity variables e, b, p, s are best thought of as deviations of the respective activity from baseline. The activity variables can hence be both positive and negative (ignoring saturation effects that arise when the baseline is very low, see below).

For our analysis, we make the following assumptions: (1) somatic input linearly increases somatic activity e, (2) dendritic input linearly increases dendritic activity b, which is in turn assumed to be independent of somatic input/activity (note that the latter assumption deviates from a BAC-firing mechanism [35], but is necessary to obtain a linear model), (3) the activities p, s of the interneuron populations increase linearly with their input, and (4) short-term plasticity is characterized by a single, static parameter (see below). Because we are interested only in qualitative statements, the analysis is done in terms of unitless variables. The model describes the dynamics of the four activity variables e, b, p, s:

$$\dot{e} = -e - w^{ep}p + E^e(t), \tag{B.1}$$

$$\dot{b} = -b - w^{bs}s + E^b(t), \tag{B.2}$$

$$\dot{p} = -p + \alpha w^{pe}e + (1-\alpha)w^{pe}b - w^{ps}s, \tag{B.3}$$

$$\dot{s} = -s + \beta w^{se}e + (1-\beta)w^{se}b - w^{sp}p. \tag{B.4}$$

Here, the synaptic weight from population y to x is modeled with a non-negative weight w^{xy} ($x, y \in \{e, p, s\}$; e: PCs, p: PV INs, s: SST INs). The central tenet of this simplified model is that somatic and dendritic activity both generate characteristic spike patterns in PCs—such as events and bursts—which are selectively transmitted by synapses because of short-term plasticity. The parameters $\alpha, \beta \in [0, 1]$ describe the short-term plasticity of the PC→PV and PC→SST synapses, respectively. $\alpha, \beta = 1$ corresponds to synapses that

only transmit somatic activity. If somatic activity generates events and dendritic activity generates bursts, this would require "perfectly depressing" synapses, i.e., synapses that transmit only the first spike of a burst. $\alpha, \beta = 0$ corresponds to synapses that only transmit dendritic activity. For the case where dendritic activity generates bursts, this requires "perfectly facilitating" synapses that ignore individual spikes and transmit only bursts. We assumed that the projections of the interneurons are specialized, i.e., that PV interneurons inhibit the soma and SST interneurons inhibit the dendrite. We will abandon this assumption in Section B.2. We also excluded inhibitory recurrence within the two populations (PV $\to$ PV, SST $\to$ SST), because these connections would only change the effective time constant of the respective activation variable. The somata and dendrites of the PCs receive time-varying external inputs $E^e(t)$ and $E^b(t)$, respectively. All activity variables follow leaky dynamics.

The dynamical system can be written as $\dot{r} = Wr + I$, where the vector r contains the activation variables $r = (e, b, p, s)^T$, I contains the external inputs $I = (E^e, E^b, 0, 0)^T$, and W is the matrix of effective connectivity strengths

$$W = \begin{pmatrix} -1 & 0 & -w^{ep} & 0 \\ 0 & -1 & 0 & -w^{bs} \\ \alpha w^{pe} & (1-\alpha)w^{pe} & -1 & -w^{ps} \\ \beta w^{se} & (1-\beta)w^{se} & -w^{sp} & -1 \end{pmatrix}. \tag{B.5}$$

Assuming that the time constant of the network is sufficiently short to adiabatically follow the input currents, we can consider the steady state by setting $\dot{r} = 0$ and solving for r:

$$Wr + I = 0 \implies r = -W^{-1}I. \tag{B.6}$$

B.1 Influence of IN→IN connections on compartment-specific E/I balance

In the steady state Eq. (B.6), the IN rates are equal to

$$p = -[W^{-1}]_{31}E^e - [W^{-1}]_{32}E^b, \tag{B.7}$$

$$s = -[W^{-1}]_{41}E^e - [W^{-1}]_{42}E^b. \tag{B.8}$$

Here, $[W^{-1}]_{ij}$ refers to the element in row i and column j of the matrix W^{-1}. Assuming that the interneurons specialize by inhibiting a single compartment, a necessary (and, up to scaling, sufficient) condition for

compartment-specific balance is that the PV rate p is proportional to the external input targeting the soma and independent of the input targeting the dendrite. Similarly, the SST rate should be proportional to the external input targeting the dendrite and independent of the input targeting the soma. By Eq. (B.7), a compartment-specific balance hence requires $[W^{-1}]_{32} = 0$ and $[W^{-1}]_{41} = 0$. Computing these matrix entries yields:

$$[W^{-1}]_{32} \propto w^{pe}(1-\alpha) - w^{ps}w^{se}(1-\beta)) = 0 \tag{B.9}$$

$$[W^{-1}]_{41} \propto -w^{pe}w^{sp}\alpha + w^{se}\beta = 0\,. \tag{B.10}$$

These equations have a simple interpretation. Each of the two terms in $[W^{-1}]_{32}$ represents a pathway by which dendritic activity reaches the PV interneurons. The first term quantifies how much dendritic activity reaches PV interneurons via the direct excitatory PC→PV projection, the second represents corresponding feedforward inhibition via the PC→SST→PV pathway. If these two pathways cancel, PV activity is independent of dendritic activity. Similarly, the two pathways in $[W^{-1}]_{41}$ that transmit somatic activity to the SST need to cancel.

What is the role of short-term plasticity? For illustration, let us first consider the limiting case of "perfect" synaptic depression ($\alpha = 1$). Perfectly depressing PC → PV synapses would imply that the PV interneurons only receive somatic activity from the PCs via the direct PC→PV pathway. The condition (B.9) then reduces to

$$[W^{-1}]_{32} \propto -w^{ps}w^{se}(1-\beta) = 0\,, \tag{B.11}$$

i.e., dendritic activity should not reach PV interneurons via the indirect PC→SST→PV pathway, because this would render PV activity dependent on dendritic activity. Because dendritic activity need to be transmitted to the SST interneurons to reach an E/I balance in the dendrite, this implies that the SST→PV connection should be absent.

"Perfect" synaptic depression ($\alpha, \beta = 1$) or facilitation ($\alpha, \beta = 0$) are hard to implement, certainly by a Markram-Tsodyks model in the presence of background activity. However, the effect of imperfect depression in the PC→PV connection ($\alpha < 1$) can be compensated by feedforward inhibition along the PC→SST→PV pathway. Similarly, imperfect PC→SST facilitation picks up somatic activity, which can then be canceled by feedforward inhibition via the PC→PV→SST pathway. The role of IN→IN synapses is therefore to complement "imperfect" short-term plasticity in decoding compartment-specific inputs.

The observation that PV→SST connections are the most important IN→IN connections in our model

results from "imperfect" facilitation in the excitatory synapses onto SST interneurons. Because events occur more frequently than bursts, the excess excitation they trigger in SST interneurons needs to be actively cancelled via the PV→SST pathway. The converse SST→PV connection is less critical, because bursts are comparatively rare, such that their transmission via PC→PV synapses causes only minor disturbances of the compartment-specific E/I balance.

B.2 Influence of IN baseline firing rates on interneuron specialization

The previous analysis assumed that interneurons were specialized to inhibit a single compartment. When should we expect specialization in the first place? We can investigate this question by extending the simplified model by inhibition from all INs onto all PC compartments:

$$W = \begin{pmatrix} -1 & 0 & -w^{ep} & -w^{es} \\ 0 & -1 & -w^{bp} & -w^{bs} \\ \alpha w^{pe} & (1-\alpha)w^{pe} & -1 & -w^{ps} \\ \beta w^{se} & (1-\beta)w^{se} & -w^{sp} & -1 \end{pmatrix}. \tag{B.12}$$

A compartment-specific balance now requires external input to be canceled by the inhibition from both interneurons:

$$E^e = w^{ep}p + w^{es}s, \tag{B.13}$$

$$E^b = w^{dp}p + w^{ds}s. \tag{B.14}$$

Without additional constraints, this system has an infinite number of solutions, i.e., weight configurations that achieve a compartment-specific balance. However, the simple constraint of low baseline firing rates of the interneurons collapses the solution space to the specialized one ($w^{es} = w^{dp} = 0$), for the following reason.

The activity variables p, s represent deviations of the interneurons firing rates from baseline. If the baseline is sufficiently high, these deviations can be both positive and negative. In that case, inhibition from one interneuron class can be cancelled by disinhibition from the other interneuron class. PV and SST interneurons are then both free to respond to both somatic and dendritic activity, as long as the weighted sums of the inhibition and disinhibition they provide to PC somata and dendrites mirrors the excitatory input to those compartments. There are many ways of doing so.

For low baseline firing rates, disinhibition is no longer available, because negative deviations from baseline are limited by the fact that activities cannot be negative. For illustration, let us consider the case where

both PV and SST neurons have zero baseline activity. By definition, the excitatory signals $E^{e/b}$ are zero for baseline activity, because they represent deviations from baseline. Therefore, we can assume that there exists a moment where the input to one PC compartment is zero, while the input to the other compartment is positive (e.g., $E^e = 0$, $E^b > 0$). By Eq. (B.13) and because all weights and rates must be positive, $w^{ep}p + w^{es}s = 0$ implies that $w^{ep} = 0$ or $p = 0$ and $w^{es} = 0$ or $s = 0$. At least one weight has to be non-zero (otherwise balancing the soma is impossible), and at least one rate has to be non-zero (otherwise balancing the dendrite is impossible). Without loss of generality we can conclude that $w^{ep} > 0, p = 0, w^{es} = 0$, and $s > 0$: The PC soma is only inhibited by the PV neuron. Analogously, the existence of a moment when $E^b = 0$ but $E^e > 0$ implies that $w^{ds} > 0, w^{dp} = 0$, meaning that the PC dendrite is only inhibited by the SST neuron. If the baseline activity is low, but not strictly zero, this saturation arguments still hold, if the variations in the firing rates that are required to balance the external input are larger than the baseline activity. Low baseline firing rates therefore imply interneuron specialization, because they prevent inhibition and disinhibition from non-specialized neurons to cancel.