

---

# Local Credit Assignment in Compartmental Dendritic Networks

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1        Backpropagation is effective but biologically implausible due to its reliance on  
2        global error transport and weight symmetry. We show that compartmental dendritic  
3        networks with shunting inhibition create a regime where strictly local learning  
4        rules can approach backpropagation-quality credit assignment. Starting from  
5        conductance-based dendritic equations, we derive exact backpropagation gradi-  
6        ents for arbitrary dendritic tree morphologies and construct a hierarchy of local  
7        approximations—3-factor (3F), 4-factor (4F), and 5-factor (5F) rules—that use  
8        only synapse-local quantities plus a broadcast error signal. Our key finding is that  
9        *shunting* (conductance-based) inhibition is critical: it provides divisive normaliza-  
10        tion that dramatically improves the alignment between local and exact gradients. A  
11        component-wise gradient-fidelity diagnostic shows that shunting networks achieve  
12         $30\times$  higher directional alignment and  $10\times$  lower scale mismatch than additive  
13        controls. The best local rule (5F with per-soma broadcast) reaches 0.914 test  
14        accuracy on MNIST, closing much of the gap to the 0.965 backpropagation ceiling  
15        on the same architecture.

## 16    1 Introduction

17    Credit assignment in deep networks is accurate with backpropagation but biologically implausible:  
18    it requires global error transport through the exact transpose of forward weights, and symmetric  
19    forward-backward pathways that have no known biological substrate. Dendritic neurons suggest  
20    an alternative architecture for learning. Real neurons possess spatially extended dendritic trees  
21    where each synapse has access to rich local state—driving forces, conductances, and branch-specific  
22    context—while global supervision may be reduced to a low-bandwidth modulatory broadcast from  
23    the soma.

24    This paper asks a concrete question: *can dendritic structure and shunting inhibition create regimes*  
25    *where strictly local learning rules approach the credit assignment quality of backpropagation?* We  
26    answer affirmatively. Starting from conductance-based dendritic equations, we derive exact backprop-  
27    agation gradients for dendritic trees and construct a hierarchy of biologically-local approximations  
28    (3-factor, 4-factor, and 5-factor rules). We then show empirically that shunting inhibition—which  
29    provides divisive normalization of compartment voltages—is the key architectural feature enabling  
30    effective local credit assignment. A novel gradient-fidelity diagnostic reveals that shunting net-  
31    works produce local gradients with  $30\times$  better directional alignment and  $10\times$  lower scale mismatch  
32    compared to additive controls.

## 33    Contributions.

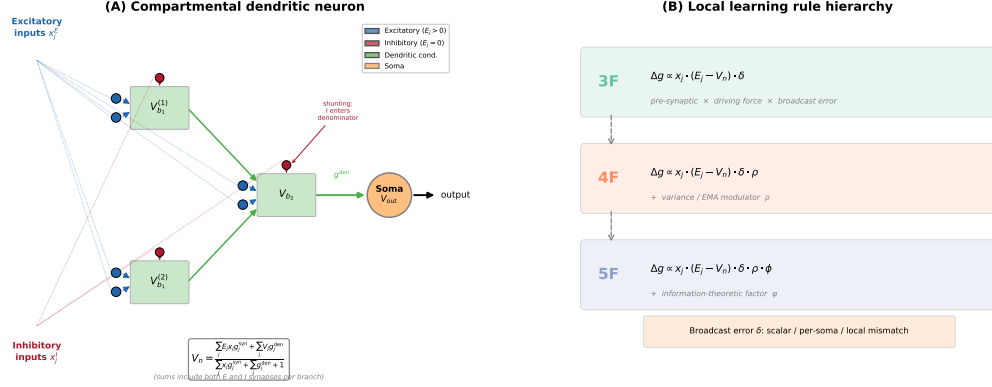


Figure 1: **Model overview.** (A) A compartmental dendritic neuron where *each branch* receives both excitatory ( $E_j > 0$ , blue) and inhibitory ( $E_j = 0$ , red) synaptic inputs via separate sparse connectivity (TopK). Inhibitory conductances enter only the denominator of the voltage equation (shunting/divisive normalization). Dendritic branch voltages propagate toward the soma via learned dendritic conductances (green). The steady-state voltage at each compartment is a conductance-weighted average (Eq. 3). (B) Local learning rules of increasing complexity: 3-factor (pre-synaptic activity  $\times$  driving force  $\times$  broadcast error), 4-factor (+ variance modulator  $\rho$ ), and 5-factor (+ information-theoretic factor  $\phi$ ). The same rule applies to both E and I synapses; the sign difference arises from the driving force ( $E_j - V_n$ ). The broadcast error  $\delta$  can operate in scalar, per-soma, or local mismatch mode.

1. **Exact dendritic backpropagation and local approximations.** We derive the exact loss gradient for arbitrary dendritic tree morphologies and construct a principled hierarchy of local approximations (3F/4F/5F) that use only quantities available at each synapse.
2. **Shunting enables local credit assignment.** We show that shunting (conductance-based) inhibition creates a regime where local rules produce gradients that are substantially better aligned with backpropagation than in additive networks, and that this translates to large performance gains.
3. **Gradient-fidelity diagnostic.** We introduce a component-wise local-vs-backprop gradient comparison that decomposes alignment by parameter type (excitatory, inhibitory, dendritic), providing mechanistic insight beyond end-task accuracy.

## 2 Compartmental Voltage Model

### 2.1 From the Passive Cable to the Compartment Equation

Starting from the linear passive cable equation for membrane potential  $v(x, t)$  relative to rest,

$$c_m \frac{\partial v}{\partial t} = \frac{1}{r_a} \frac{\partial^2 v}{\partial x^2} - \frac{1}{r_m} v + i_{\text{syn}}(x, t),$$

and discretizing a dendritic branch into isopotential compartments with axial conductances, the steady-state ( $\partial_t v = 0$ ) yields a nodal balance of conductances and driving forces.<sup>1</sup> Representing synapses as conductances with reversal potentials and siblings as dendritic conductances gives the compartment equation below with unit leak to  $E_{\text{leak}} = 0$ . This clarifies that (i) all inputs contribute via conductances, (ii) total conductance  $g_n^{\text{tot}}$  controls both input resistance  $R_n^{\text{tot}}$  and divisive normalization, and (iii) shunting inhibition corresponds to adding conductance with  $E_{\text{inh}} \approx 0$  (Section 2.4).

### 2.2 Voltage Equation

Consider compartment  $n$  receiving synaptic inputs indexed by  $j$  and dendritic inputs from child compartments. Let:

<sup>1</sup>See Koch's *Biophysics of Computation* and Dayan & Abbott's *Theoretical Neuroscience* for derivations and assumptions underlying the linear regime.

- 56 •  $x_j \in \mathbb{R}_+$ : presynaptic activity at synapse  $j$
- 57 •  $E_j \in \mathbb{R}$ : reversal potential of synapse  $j$  (excitatory:  $E_j > 0$ ; inhibitory:  $E_j \leq 0$ )
- 58 •  $g_j^{\text{syn}} \geq 0$ : synaptic conductance (learned parameter)
- 59 •  $V_j \in \mathbb{R}$ : voltage of child compartment  $j$
- 60 •  $g_j^{\text{den}} \geq 0$ : dendritic conductance from child  $j$  (learned parameter)

61 **Currents.** Synaptic current:

$$I_{\text{syn}} = \sum_j (E_j - V_n) x_j g_j^{\text{syn}} \quad (1)$$

62 Dendritic current:

$$I_{\text{den}} = \sum_j (V_j - V_n) g_j^{\text{den}} \quad (2)$$

63 **Steady-state voltage.** With unit leak conductance to reversal potential 0:

$$V_n = \frac{\sum_j E_j x_j g_j^{\text{syn}} + \sum_j V_j g_j^{\text{den}}}{\sum_j x_j g_j^{\text{syn}} + \sum_j g_j^{\text{den}} + 1} \quad (3)$$

**Total conductance and resistance.**

$$g_n^{\text{tot}} = \sum_j x_j g_j^{\text{syn}} + \sum_j g_j^{\text{den}} + 1, \quad R_n^{\text{tot}} = \frac{1}{g_n^{\text{tot}}} \quad (4)$$

64 **Lemma 1** (Convexity and Bounds). *Let  $\mathcal{S}_n = \{E_j\}_{\text{syn at } n} \cup \{V_j\}_{\text{children}} \cup \{0\}$ . Then  $V_n$  in (3) is a*  
 65 *convex combination of elements of  $\mathcal{S}_n$ , hence*

$$\min \mathcal{S}_n \leq V_n \leq \max \mathcal{S}_n.$$

66 *Moreover,  $0 < R_n^{\text{tot}} \leq 1$  and  $R_n^{\text{tot}} g_i^{\text{den}} < 1$  for all  $i$ .*

67 *Proof.* Immediate from (3)(4) since all conductances are nonnegative and leak adds +1 to the  
 68 denominator.  $\square$

## 69 2.3 Local Sensitivities

**Proposition 1** (Synaptic Gradient).

$$\frac{\partial V_n}{\partial g_i^{\text{syn}}} = x_i R_n^{\text{tot}} (E_i - V_n) \quad (5)$$

70 *Proof.* Apply quotient rule to (3):

$$\begin{aligned} \frac{\partial V_n}{\partial g_i^{\text{syn}}} &= \frac{E_i x_i \cdot g_n^{\text{tot}} - \left( \sum_j E_j x_j g_j^{\text{syn}} + \sum_j V_j g_j^{\text{den}} \right) \cdot x_i}{(g_n^{\text{tot}})^2} \\ &= \frac{E_i x_i}{g_n^{\text{tot}}} - \frac{V_n x_i}{g_n^{\text{tot}}} = x_i R_n^{\text{tot}} (E_i - V_n). \end{aligned} \quad \square$$

**Proposition 2** (Child Voltage Gradient).

$$\frac{\partial V_n}{\partial V_i} = g_i^{\text{den}} R_n^{\text{tot}} \quad (6)$$

**Proposition 3** (Dendritic Gradient).

$$\frac{\partial V_n}{\partial g_i^{\text{den}}} = R_n^{\text{tot}} (V_i - V_n) \quad (7)$$

**Proposition 4** (Additional Local Sensitivities).

$$\frac{\partial V_n}{\partial x_i} = g_i^{\text{syn}} R_n^{\text{tot}} (E_i - V_n), \quad \frac{\partial V_n}{\partial E_i} = x_i g_i^{\text{syn}} R_n^{\text{tot}}, \quad \frac{\partial V_n}{\partial g^{\text{leak}}} = -V_n R_n^{\text{tot}}.$$

71 **Remark 1.** *The sensitivity  $\partial V_n / \partial g^{\text{leak}} = -V_n R_n^{\text{tot}}$  applies if  $g^{\text{leak}}$  is a trainable parameter. In all*  
 72 *reported experiments, we fix  $g^{\text{leak}}=1$  (normalized units).*

## 73 2.4 Shunting Inhibition and Divisive Gain Control

74 A conductance-based inhibitory synapse with  $E_{\text{inh}} \approx E_{\text{leak}}=0$  contributes current  $I_{\text{inh}} = (0 -$   
75  $V_n) x_j g_j^{\text{syn}}$  and increases  $g_n^{\text{tot}}$  in (4).

76 **Proposition 5** (Subthreshold Effect of Shunts). *For a pure shunt ( $E_j = 0$ ), the steady-state sensitivity*  
77 *to the inhibitory conductance is*

$$\frac{\partial V_n}{\partial g_j^{\text{syn}}} = x_j R_n^{\text{tot}} (0 - V_n) = -x_j R_n^{\text{tot}} V_n.$$

78 Thus  $V_n$  is multiplicatively attenuated (divisive normalization) by increased inhibitory conductance  
79 at fixed drives.

80 **Remark 2** (Divisive vs. Subtractive at the Firing-Rate Level). *While shunting produces divisive*  
81 *scaling of subthreshold voltages, its net effect on firing rates can be subtractive in many regimes*  
82 *[4], so we report both voltage- and rate-level analyses in experiments. Normalization via added*  
83 *conductance is consistent with canonical divisive normalization models in cortex [3].*

84 **Inhibitory/shunting synapses.** For an inhibitory synapse with  $E_j \approx 0$ , the 3F update reduces to

$$\Delta g_{j,\text{inh}}^{\text{syn}} = \eta \langle x_j R_n^{\text{tot}} (-V_n) e_n \rangle_B,$$

85 i.e., anti-Hebbian in  $V_n$  and divisive in  $g_n^{\text{tot}}$ . With 4F/5F, multiply by  $\rho$  and  $\phi$  (Def. 4). Note  
86 that the same multiplicative factors are applied to both excitatory and inhibitory synapses in the  
87 implementation; the sign difference arises solely from the driving force ( $E_j - V_n$ ).

## 88 2.5 Loss Propagation

89 Let  $V_0$  denote the somatic/output compartment. The decoder produces  $\hat{y} = W_{\text{dec}} V_0$  (linear case), and  
90  $L$  is the task loss. Define the error gradients:

$$\delta^y := \frac{\partial L}{\partial \hat{y}}, \quad \delta_0 := \frac{\partial L}{\partial V_0} = \left( \frac{\partial \hat{y}}{\partial V_0} \right)^\top \delta^y = W_{\text{dec}}^\top \delta^y. \quad (8)$$

91 **Theorem 1** (Backpropagation on a Dendritic Tree). *Let the dendritic morphology be a rooted tree*  
92 *with soma/output at node 0. For any compartment  $n$  with parent set  $\mathcal{P}(n)$  (typically  $|\mathcal{P}(n)|=1$ ), the*  
93 *loss gradient satisfies the recursion*

$$\frac{\partial L}{\partial V_n} = \sum_{p \in \mathcal{P}(n)} \frac{\partial L}{\partial V_p} \frac{\partial V_p}{\partial V_n} = \sum_{p \in \mathcal{P}(n)} \delta_p R_p^{\text{tot}} g_{n \rightarrow p}^{\text{den}}, \quad \delta_p \equiv \frac{\partial L}{\partial V_p}. \quad (9)$$

94 Unrolling the recursion yields a sum over all directed paths  $\mathcal{P} : n \rightsquigarrow 0$ :

$$\frac{\partial L}{\partial V_n} = \frac{\partial L}{\partial V_0} \sum_{\mathcal{P}: n \rightsquigarrow 0} \prod_{(i \rightarrow k) \in \mathcal{P}} R_k^{\text{tot}} g_{i \rightarrow k}^{\text{den}}. \quad (10)$$

95 *Proof.* Apply the multivariate chain rule on the directed acyclic computation graph defined by the  
96 tree; use Proposition 2. Each path contributes a product of edge sensitivities. Summing over parents  
97 produces (9); unrolling yields (10).  $\square$

98 **Corollary 1** (Chain Case). *If the morphology is a simple chain  $V_0 \leftarrow V_1 \leftarrow \dots \leftarrow V_n$ , (10) reduces*  
99 *to*

$$\frac{\partial L}{\partial V_n} = \frac{\partial L}{\partial V_0} \prod_{i=1}^n R_{i-1}^{\text{tot}} g_i^{\text{den}}. \quad (11)$$

100 Defining  $g_0^{\text{den}} = 1$  and reindexing:

$$\frac{\partial L}{\partial V_n} = \frac{\partial L}{\partial V_0} \prod_{i=0}^n R_i^{\text{tot}} g_i^{\text{den}}. \quad (12)$$

**Corollary 2** (Synaptic Parameter Gradient).

$$\frac{\partial L}{\partial g_j^{\text{syn}}} = \frac{\partial L}{\partial V_0} \left( \prod_{i=0}^n R_i^{\text{tot}} g_i^{\text{den}} \right) x_j (E_j - V_n) \quad (13)$$

**Corollary 3** (Dendritic Parameter Gradient).

$$\frac{\partial L}{\partial g_j^{\text{den}}} = \frac{\partial L}{\partial V_0} \left( \prod_{i=0}^n R_i^{\text{tot}} g_i^{\text{den}} \right) (V_j - V_n) \quad (14)$$

### 101 3 Local Learning Approximations

#### 102 3.1 Broadcast Error Approximation

103 **Definition 1** (Local Approximation). *Replace the exact gradient  $\frac{\partial L}{\partial V_n}$  with a broadcast error signal*  
 104  *$e_n$  derived from the output error  $\delta_0 = \frac{\partial L}{\partial V_0}$ :*

$$\frac{\partial L}{\partial V_n} \approx e_n, \quad \prod_{i=0}^n R_i^{\text{tot}} g_i^{\text{den}} \approx 1 \quad (15)$$

105 Three broadcast modes are considered:

106 **(A) Scalar broadcast.** For minibatch index  $b$ :

$$\bar{\delta}(b) = \frac{1}{d_{\text{out}}} \sum_{k=1}^{d_{\text{out}}} \delta_k(b), \quad e_n(b) = \bar{\delta}(b) \mathbf{1}_{d_n} \quad (16)$$

107 **(B) Per-compartment mapping.** If  $d_n = d_{\text{out}}$ :  $e_n(b) = \delta(b)$ . Otherwise, *fallback to scalar*  
 108 *broadcast.* An optional DFA-style mode uses a fixed random feedback matrix  $B_n \in \mathbb{R}^{d_n \times d_{\text{out}}}$   
 109 sampled once at initialization:  $e_n(b) = B_n \delta(b)$ . This supports testing Theorem 2.

110 **(C) Local mismatch modulation.** Let  $P_n(b)$  be parent compartment drive (e.g., blocklinear output).  
 111 Define centered mismatch:

$$\varepsilon_n(b) = (P_n(b) - V_n(b)) - \frac{1}{B} \sum_{t=1}^B (P_n(t) - V_n(t)) \quad (17)$$

112 Then:

$$e_n(b) = \bar{\delta}(b) \varepsilon_n(b) \quad (18)$$

#### 113 3.2 Gradient Alignment with Broadcast Errors

114 Define the exact synaptic gradient at layer  $n$  by  $g^{\text{exact}} = \delta_0 \cdot \Xi_n$ , where  $\Xi_n$  collects local factors and  
 115 the exact path-sum (10). The local 3F gradient with broadcast error  $e_n = B_n \delta_0$  is  $g^{\text{local}} = e_n \cdot \hat{\Xi}_n$ ,  
 116 where  $\hat{\Xi}_n$  omits the path-sum.

117 **Theorem 2** (Positive Expected Alignment under Random Broadcast). *Let  $B_n \in \mathbb{R}^{d_n \times d_{\text{out}}}$  have i.i.d.*  
 118 *zero-mean entries with  $\mathbb{E}[B_n^\top B_n] = \alpha I$ . If the decoder aligns with the forward pathway (standard*  
 119 *during training), then*

$$\mathbb{E}[\cos \angle(g^{\text{local}}, g^{\text{exact}})] \geq c_n > 0,$$

120 *where  $c_n$  depends on  $\alpha$  and the average correlation between  $\hat{\Xi}_n$  and  $\Xi_n$ . Thus  $g^{\text{local}}$  provides a*  
 121 *descent direction in expectation.*

122 *Sketch.* Adapt the feedback-alignment argument [5, 6]: fixed random feedback suffices for alignment  
 123 as forward weights adapt. Here,  $\hat{\Xi}_n$  is proportional to  $\Xi_n$  up to the missing path factor; Jensen bounds  
 124 on (10) yield  $c_n > 0$ .  $\square$

### 125 3.3 Three-Factor Rule (3F)

126 **Definition 2** (3F Learning Rule). *For synaptic conductances:*

$$\Delta g_j^{\text{syn}} = \eta \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B \quad (19)$$

127 *For dendritic conductances:*

$$\Delta g_j^{\text{den}} = \eta \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B \quad (20)$$

128 *where  $\langle \cdot \rangle_B$  denotes batch average.*

129 **Remark 3.** *The three factors are: (1) presynaptic activity  $x_j$  or voltage difference  $(V_j - V_n)$ , (2)*  
 130 *postsynaptic modulation  $(E_j - V_n)$  or  $R_n^{\text{tot}}$ , (3) broadcast error  $e_n$ .*

131 **Symmetry note.** *The same multiplicative factors ( $R_n^{\text{tot}}$ ,  $\rho$ ,  $\phi$ ,  $s_j$ ) apply to both excitatory and*  
 132 *inhibitory synapses; the sign difference arises solely from the driving force  $(E_j - V_n)$ .*

### 133 3.4 Four-Factor Rule (4F): Morphology Correlation

134 **Definition 3** (Morphology Factor). *Let  $\bar{V}_n = \frac{1}{d_n} \sum_{j=1}^{d_n} V_{n,j}$  be the mean voltage over compartments*  
 135 *in layer  $n$ . Define the correlation with output:*

$$\rho_n = \frac{\text{Cov}(\bar{V}_n, \bar{V}_0)}{\sqrt{\text{Var}(\bar{V}_n) \text{Var}(\bar{V}_0) + \varepsilon}} \quad (21)$$

136 **Proposition 6** (4F Update Rule). *Multiply 3F updates by  $\rho_n$ :*

$$\Delta g_j^{\text{syn}} = \eta \rho_n \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B \quad (22)$$

$$\Delta g_j^{\text{den}} = \eta \rho_n \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B \quad (23)$$

137 **Proposition 7** (Approximate Gradient Alignment). *Let  $L$  be a smooth loss. If layer  $n$  contributes to*  
 138 *the output primarily through its mean activity, then  $\mathbb{E}[\frac{\partial L}{\partial \bar{V}_n} \cdot \bar{V}_n] \propto \rho_n \cdot \text{Var}(\bar{V}_n)$ . Thus  $\rho_n$  weights*  
 139 *updates by the layer's relevance to the task.*

140 **Estimators (EMA / online).** For minibatches  $B \geq 2$ , estimate  $\rho_n$  from (21) with an EMA over  
 141 batches. For  $B = 1$  (online), maintain means  $\mu_x, \mu_y$ , variances  $\sigma_x^2, \sigma_y^2$ , and covariance  $C_{xy}$  for  
 142  $x_t = \bar{V}_0^{(t)}$  and  $y_t = \bar{V}_n^{(t)}$  using Welford's numerically stable algorithm [16]:

$$\begin{aligned} \mu_x^{(t)} &= (1 - \alpha) \mu_x^{(t-1)} + \alpha x_t, & \mu_y^{(t)} &= (1 - \alpha) \mu_y^{(t-1)} + \alpha y_t, \\ \delta_x &= x_t - \mu_x^{(t-1)}, & \delta_y &= y_t - \mu_y^{(t-1)}, \\ \sigma_x^{2(t)} &= (1 - \alpha) \sigma_x^{2(t-1)} + \alpha \delta_x^2, & \sigma_y^{2(t)} &= (1 - \alpha) \sigma_y^{2(t-1)} + \alpha \delta_y^2, \\ C_{xy}^{(t)} &= (1 - \alpha) C_{xy}^{(t-1)} + \alpha \delta_x \delta_y. \end{aligned} \quad (24)$$

143 Then  $\rho_n^{(t)} = C_{xy}^{(t)} / (\sqrt{\sigma_x^{2(t)} \sigma_y^{2(t)}} + \varepsilon)$ , where  $\alpha$  is the EMA rate.

### 144 3.5 Five-Factor Rule (5F): Conditional Information

145 **Definition 4** (Conditional Predictability Factor). *Let  $P_n$  be parent compartment voltage. Define the*  
 146 *conditional variance via ridge regression:*

$$\beta_n = \frac{\text{Cov}(V_n, P_n)}{\text{Var}(P_n) + \lambda} \quad (25)$$

$$\sigma_{\text{res}}^2 = \text{Var}(V_n) - \beta_n \text{Cov}(V_n, P_n) \quad (26)$$

147 *The information proxy is:*

$$\phi_n = \frac{\text{Var}(V_n)}{\sigma_{\text{res}}^2 + \varepsilon} = \frac{1}{1 - R_n^2} \geq 1, \quad (27)$$

148 *where  $R_n^2 = \frac{\beta_n \text{Cov}(V_n, P_n)}{\text{Var}(V_n)}$  is the (ridge) coefficient of determination.*

**Remark 4** (Information-Theoretic Interpretation).  $\phi_n$  increases when  $V_n$  is more predictable from its parent  $P_n$  (higher  $R^2$ ), amplifying updates for compartments with strong signal propagation. In practice,  $\phi_n$  is clamped to  $[0.25, 4.0]$  for stability. An alternative  $\phi_n = 1 - R^2$  would instead emphasize compartments with unique information; both are valid depending on whether coherent signal flow or novelty is prioritized.

**Proposition 8** (5F Update Rule).

$$\Delta g_j^{\text{syn}} = \eta \rho_n \phi_n \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B \quad (28)$$

$$\Delta g_j^{\text{den}} = \eta \rho_n \phi_n \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B \quad (29)$$

## 4 Morphology-Aware Extensions

Standard 4F/5F rules use layer-wise factors  $\rho_n, \phi_n$  that ignore branch-specific topology. We introduce four extensions that explicitly incorporate dendritic tree structure.

### 4.1 Path-Integrated Propagation

**Definition 5** (Path Factor). Define recursively:

$$\pi_n = \begin{cases} 1 & n = 0 \\ \pi_{n-1} \cdot R_{n-1}^{\text{tot}} \cdot \bar{g}_{n-1}^{\text{den}} & n \geq 1 \end{cases} \quad (30)$$

where  $\bar{g}_{n-1}^{\text{den}}$  is the mean dendritic conductance from layer  $n - 1$  to  $n$ .

**Proposition 9** (Exact for Chains). For a chain morphology (single path), the path factor (30) satisfies

$$\pi_n = \prod_{i=0}^{n-1} R_i^{\text{tot}} g_i^{\text{den}} \quad (31)$$

and thus exactly matches (10).

**Remark 5** (Sandwich bounds for trees). For a tree, let  $\mathcal{P}$  be the set of directed paths from  $n$  to 0 and define, at each depth  $d$ ,  $m_d := \min_{\mathcal{P}} a_{d,\mathcal{P}}$  and  $M_d := \max_{\mathcal{P}} a_{d,\mathcal{P}}$ , where  $a_{d,\mathcal{P}}$  is the edge factor  $R_k^{\text{tot}} g_{i \rightarrow k}^{\text{den}}$  at depth  $d$  along path  $\mathcal{P}$ . Then

$$|\mathcal{P}| \prod_d m_d \leq \sum_{\mathcal{P}} \prod_d a_{d,\mathcal{P}} \leq |\mathcal{P}| \prod_d M_d.$$

If per-depth factors are narrowly concentrated ( $m_d \approx M_d$ ), replacing the sum by a product of per-depth means (our  $\pi_n$ ) is accurate. Modulating the error by  $\pi_n$  yields  $e_n^{\text{path}} = e_n \cdot \pi_n$ , which better approximates exact backpropagation.

**Corollary 4** (Depth Attenuation). From Lemma 1,  $R_k^{\text{tot}} g_{i \rightarrow k}^{\text{den}} < 1$ . Therefore any product  $\prod R g$  in (10) decays exponentially with depth, motivating path-based error attenuation.

**Remark 6.** In practice,  $\pi_n$  is computed as a per-sample scalar (broadcast to all compartments in layer  $n$ ), using the arithmetic mean of outgoing dendritic conductances at each depth. The modulated error is  $e_n \leftarrow e_n \cdot \pi_n$ .

**Theoretical effect.** Path propagation introduces depth-dependent attenuation: deeper compartments receive exponentially smaller error signals  $e_n \sim \prod R_i g_i$ . This encourages specialization: shallow layers learn direct input-output mappings, while deep layers integrate over longer paths. In practice, we use a per-sample scalar path factor  $\pi_n(b)$  for stability and consistent broadcasting across layers with different widths.

### 4.2 Additional Extensions

Beyond path propagation, we implement three further morphology-aware mechanisms (detailed in Appendix A):

- **Depth modulation:** Per-branch scaling  $\rho_j = \rho_{\text{base}} / (d_j + \alpha)$  that mirrors cable attenuation, biasing learning toward proximal synapses.

183 • **Dendritic normalization:** Update normalization by total branch conductance,  $\Delta g_j^{\text{den}} \leftarrow$   
 184  $\Delta g_j^{\text{den}} / (\sum_k g_k^{\text{den}} + \varepsilon)$ , which stabilizes update variance analogous to homeostatic scaling  
 185 [17].

186 • **Apical/basal differentiation:** Branch-type-specific scaling factors  $s_j$  that allow differential  
 187 plasticity in feedback (apical) vs. feedforward (basal) compartments [15].

188 **HSIC auxiliary objectives.** We optionally add Hilbert-Schmidt Independence Criterion (HSIC)  
 189 losses [11] as auxiliary objectives: a self-decorrelation term that encourages diverse representations  
 190 within each layer, and a target-correlation term that aligns layer activations with labels. Moderate  
 191 HSIC weights improve performance on context-gating tasks while having negligible effect on MNIST  
 192 (see Section 7). Full definitions and gradients are in Appendix B.

## 193 5 Implementation Details

### 194 5.1 Units and Normalization

Quantity	Symbol	Typical units (scaled)
Voltage	$V$	mV (normalized to $[-1, 1]$ )
Synaptic conductance	$g^{\text{syn}}$	nS (nonnegative)
Dendritic conductance	$g^{\text{den}}$	nS (nonnegative)
Leak conductance	$g^{\text{leak}}$	nS (set to 1 in normalized units)
Input resistance	$R^{\text{tot}}$	$\text{nS}^{-1}$ (normalized $\leq 1$ )

Table 1: Units and normalization conventions.

### 195 5.2 Positive Weight Parameterization

196 To enforce  $g \geq 0$ , we can use exponential:

$$g = \exp(\theta), \quad \theta \in \mathbb{R} \quad (32)$$

197 Chain rule for gradients:

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial g} \cdot g \quad (33)$$

198 Alternatively, use softplus  $g = \log(1 + \exp(\theta))$  to avoid extreme gradients.

### 199 5.3 Decoder Update Modes

200 Let  $W_{\text{dec}}$  map  $V_L \rightarrow y \in \mathbb{R}^{d_{\text{out}}}$ . Three modes:

- 201 1. **Backprop:**  $\nabla_{W_{\text{dec}}} L$  via autograd.
- 202 2. **Local:**  $\Delta W_{\text{dec}} = \eta \langle \delta_0 V_L^\top \rangle_B$  (3-factor).
- 203 3. **Frozen:**  $\Delta W_{\text{dec}} = 0$ .



**Algorithm 1** Local Credit Assignment with Morphology-Aware Extensions

---

```

1: Input: Model, minibatch  $(x, y)$ , config  $\mathcal{C}$ 
2: Forward pass:  $\hat{y} = f(x; \mathbf{g}^{\text{syn}}, \mathbf{g}^{\text{den}})$ 
3: Compute loss  $L$  and output error  $\delta^y = \frac{\partial L}{\partial \hat{y}}$ 
4: Compute somatic error  $\delta_0 = W_{\text{dec}}^\top \delta^y$ 
5: for each layer  $n$  (reverse order) do
6:   Broadcast error:  $e_n = \text{broadcast}(\delta_0, \mathcal{C})$ 
7:   if path propagation enabled then
8:     Compute  $\pi_n$  via (30);  $e_n \leftarrow e_n \cdot \pi_n$ 
9:   end if
10:  Compute  $\rho_n$  via (21)
11:  if depth modulation enabled then
12:     $\rho_n \leftarrow [\rho_1, \dots, \rho_{d_n}]$  via (??)
13:  end if
14:  Compute  $\phi_n$  via (27)
15:  Compute branch scales  $s_j$  via (??)
16:  Synaptic updates:  $\Delta g_j^{\text{syn}} = \eta s_j \rho_j \phi_n \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B$ 
17:  Dendritic updates:  $\Delta g_j^{\text{den}} = \eta s_j \rho_j \phi_n \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B$ 
18:  if dendritic normalization enabled then
19:     $\Delta g_j^{\text{den}} \leftarrow \Delta g_j^{\text{den}} / (G_n + \varepsilon)$  via (??)
20:  end if
21:  if HSIC enabled then
22:    Add HSIC gradients to  $\Delta g_j^{\text{syn}}$ 
23:  end if
24: end for
25: Clip gradients; optimizer step

```

---

205 **6 Theoretical Comparison**

Method	Factors	Complexity	Best observed regime (current sweeps)
3F	$x, (E - V), e$	$\mathcal{O}(1)$	Baseline local plasticity; weak on contextual/hierarchical tasks
4F	$3F + \rho$	$\mathcal{O}(1)$	Better scaling/conditioning than 3F; still limited performance ceiling
5F	$4F + \phi$	$\mathcal{O}(d_n)$	Strongest overall local competence (MNIST, context gating)
5F + Path	$5F + \pi$	$\mathcal{O}(L)$	Strongest impact on representation metrics; selective accuracy gains
5F + Depth	$5F, \rho \rightarrow \rho_j$	$\mathcal{O}(d_n)$	Useful in deeper/branched morphologies
5F + Norm	$5F + \text{normalization}$	$\mathcal{O}(d_n)$	Stabilizes update scale across branches
5F + Types	$5F \times s_j$	$\mathcal{O}(1)$	Tests apical/basal specialization hypotheses

Table 2: Variant taxonomy: computational cost and current empirical regime map ( $L = \text{depth}$ ,  $d_n = \text{compartments}$ ).

206 **7 Experiments**207 **7.1 Setup and Main Claims**

208 We evaluate local credit assignment rules through a structured empirical program that separates  
209 (i) model capacity, (ii) local-rule competence, and (iii) mechanistic regime dependence. Our main  
210 findings are:

- 211 **1. Shunting is the key regime:** Shunting dendritic cores outperform additive controls under  
212 the same local rule, with the gap widening under stronger inhibition (Figs. 12, 11).

Component	Biological Analog	Key Result
Conductance scaling $R_n^{\text{tot}}$	Input resistance	Lemma 1: $0 < R_n^{\text{tot}} \leq 1$
Driving force ( $E_j - V_n$ )	Synaptic current	Prop. 1: Local sensitivity
Shunting inhibition	Divisive normalization	Sec. 2.4: $\partial V / \partial g_{\text{inh}} \propto -V$
Path factor $\pi_n$	Cable attenuation	Prop. 9: Depth decay
Morphology factor $\rho_n$	Layer correlation	Eq. (21): Task relevance
Information factor $\phi_n$	Conditional predictability	Eq. (27): $1/(1 - R^2)$
Dendritic normalization	Homeostatic scaling	Sec. 4.3: Variance stabilization
Branch-type scaling	Apical vs. basal	Sec. 4.4: Compartment specialization
Broadcast alignment	Feedback alignment	Thm. 2: $\mathbb{E}[\cos \angle] > 0$

Table 3: Summary of theoretical components and their biological/algorithmic interpretations.

Dataset	Rule	Top-10 valid mean	Top-10 test mean
MNIST	3F	0.611	0.622
MNIST	4F	0.620	0.628
MNIST	5F	<b>0.912</b>	<b>0.916</b>
Context gating	3F	0.398	0.396
Context gating	4F	0.411	0.411
Context gating	5F	<b>0.807</b>	<b>0.789</b>

Table 4: **Rule-family ranking from completed local-competence sweeps.** Values are averages over the top 10 runs (ranked by validation accuracy) within each dataset and rule family. 5F dominates 3F/4F in both datasets under the current sweep space.

2. **Gradient fidelity explains the advantage:** Shunting models produce local gradients with dramatically better alignment to exact backpropagation (Table 5, Fig. 5).
3. **5F with per-soma broadcast is the best local rule:** It reaches 0.914 test accuracy on MNIST (0.965 backprop ceiling) and 0.803 on context gating (0.864 ceiling).

## 7.2 Phase-Based Evaluation (Phase 1–Phase 3)

We use a phase-based sweep suite:

- **Phase 1 (capacity ceiling):** train identical architectures with standard backprop to establish a meaningful accuracy ceiling per dataset and core type (Fig. 8).
- **Phase 2 (local competence):** within the Phase-1 capacity regime, sweep LocalCA knobs (rule variant, broadcast mode, morphology-aware modulators, HSIC auxiliaries) to find competent local configurations.
- **Phase 3 (mechanistic claims):** targeted sweeps testing regime dependence (shunting vs additive across inhibition/noise), morphology scaling, error shaping, and information analyses.

## 7.3 Variant Ranking Across Local Rule Families

To make the rule comparison explicit, we aggregate completed Phase-2 and Phase-2b runs on MNIST and context gating and rank 3F/4F/5F under matched local-learning sweeps.

**Broadcast-mode interaction.** Per-soma broadcast strongly outperforms scalar and local-mismatch modes for both datasets (MNIST: 0.918 vs 0.889 vs 0.200 test; context gating: 0.827 vs 0.699 vs 0.114 test). This motivates treating the broadcast mode as a primary experimental factor.

## 7.4 Results

**Capacity ceilings (standard backprop).** Standard backprop achieves 0.978 test accuracy on MNIST (point MLP) and 0.965 (dendritic shunting), 0.864 on context gating (dendritic shunting), and  $\approx 0.49$  on CIFAR-10 with flattened inputs (Fig. 8). These ceilings define the reference for local-rule comparisons.

## MNIST: Training Strategy Comparison (Matched Architecture)

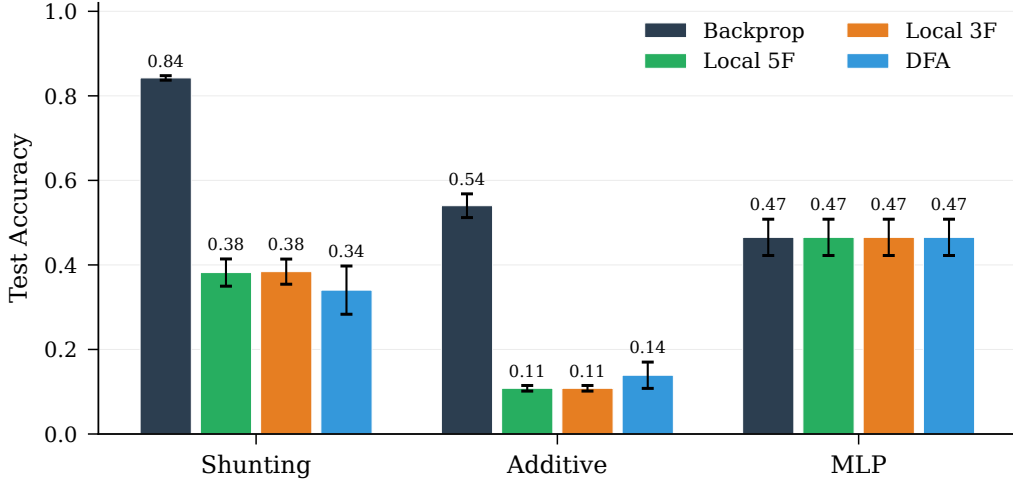


Figure 2: **Training strategy comparison on MNIST (matched architecture).** Test accuracy for backpropagation, local rules (5F, 3F with per-soma broadcast), and DFA across three architectures. Shunting dendritic cores provide the strongest backprop ceiling (0.84) and the largest local-rule advantage over additive controls. DFA performs comparably to local rules on dendritic architectures but matches backprop on the point MLP. Error bars: 95% CI.

**Local competence.** With the best local configuration (5F rule, per-soma broadcast, dendritic shunting with layer size 128, branch factors [3, 3], 20 inhibitory synapses), local credit assignment reaches  $0.914 \pm 0.003$  test accuracy on MNIST and  $0.803 \pm 0.006$  on context gating (with moderate HSIC weight), compared to backprop ceilings of 0.965 and 0.864 respectively (Fig. 9).

**Regime dependence.** The shunting advantage is largest in high-inhibition regimes: per-soma broadcast with shunting shows  $\approx 0.19$  test-accuracy advantage over additive controls at higher inhibition levels. Representation-level metrics (mutual information) show clear broadcast/path interactions (Fig. 14).

**Baseline comparison on matched architectures.** To contextualize local learning performance against alternative biologically plausible algorithms, we compare backpropagation, local rules (5F, 3F), and direct feedback alignment (DFA) on MNIST using matched small architectures (Fig. 2). Backpropagation on the shunting dendritic core achieves 0.84 test accuracy, while local 5F and 3F rules reach  $\approx 0.38$ —substantially above the additive control ( $\approx 0.11$ ), confirming shunting as the key enabler. DFA achieves 0.34 on the shunting core but matches backpropagation on point MLP (0.47), consistent with its known strength in sequential architectures. Notably, feedback alignment (FA) fails entirely on dendritic architectures due to their non-sequential topology while DFA works across all architectures.

**Learning dynamics.** Figure 3 shows learning curves across training strategies. Backpropagation converges fastest, with the shunting dendritic core reaching high accuracy within 10 epochs. Local rules converge more slowly but steadily improve throughout training. On Fashion-MNIST, local 5F achieves the strongest performance on shunting cores ( $\approx 0.36$ ), maintaining the shunting advantage observed on MNIST.

**Fashion-MNIST generalization.** Local rules transfer to Fashion-MNIST (Fig. 4), a substantially harder variant. With per-soma broadcast, shunting cores reach  $\approx 0.35$  test accuracy across all three rule variants (5F: 0.358, 4F: 0.352, 3F: 0.344), while additive controls collapse to chance ( $\approx 0.10$ ). The shunting advantage ( $\approx 3.5\times$ ) is even larger than on MNIST, suggesting that shunting-based local credit assignment becomes *more* important as task difficulty increases.

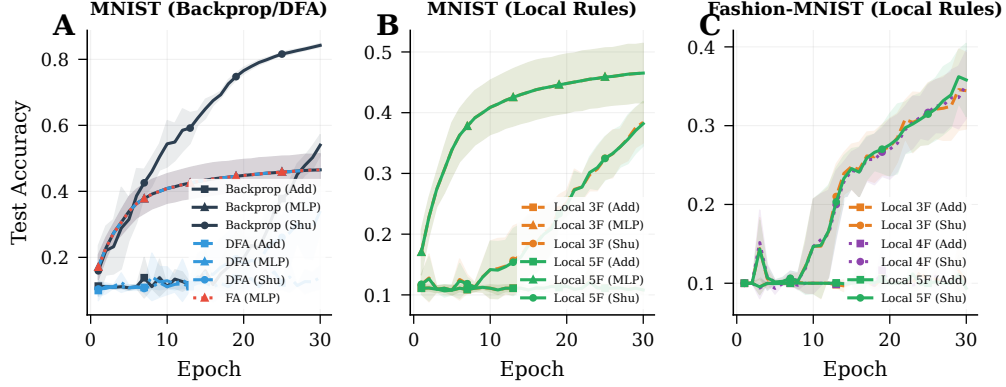


Figure 3: **Learning curves across training strategies.** (A) MNIST with backpropagation and DFA baselines, showing faster convergence for backprop on shunting cores. (B) MNIST with local rules (3F/4F/5F), comparing shunting, additive, and MLP architectures. (C) Fashion-MNIST with local rules, confirming the shunting advantage generalizes to a harder dataset. Shaded regions:  $\pm 1$  s.d. across seeds.

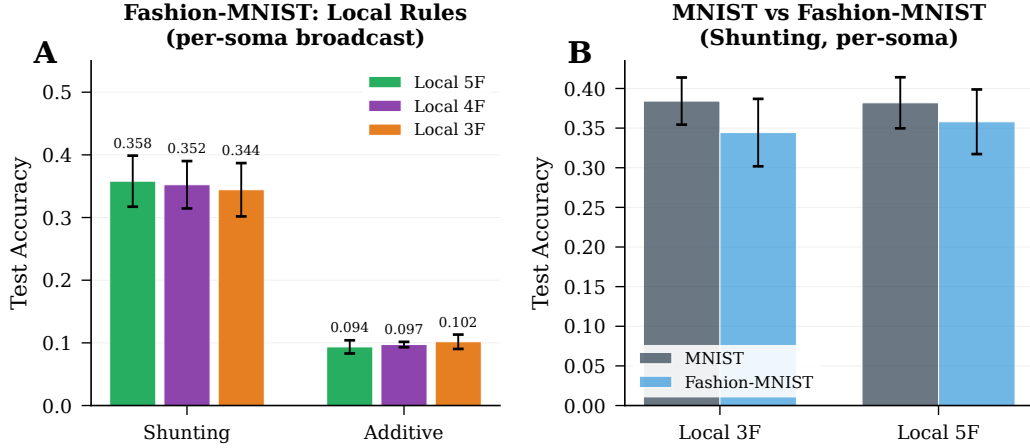


Figure 4: **Fashion-MNIST local learning results.** (A) Test accuracy by rule variant and architecture (per-soma broadcast). Shunting cores outperform additive controls by  $\approx 3.5\times$ . (B) MNIST vs Fashion-MNIST comparison on shunting cores with per-soma broadcast. Both datasets show similar absolute performance under local rules, with Fashion-MNIST marginally lower. Error bars: 95% CI.

## 264 7.5 Gradient-Fidelity Diagnostic (Local vs Backprop)

265 To test whether improved performance corresponds to better credit signals, we compare LocalCA and  
 266 backprop gradients on the *same batch and same initial weights*, component-wise. For each parameter  
 267 tensor  $p$ , we compute

$$\cos_p = \frac{\langle g_p^{\text{local}}, g_p^{\text{bp}} \rangle}{\|g_p^{\text{local}}\|_2 \|g_p^{\text{bp}}\|_2 + \varepsilon}, \quad (34)$$

268 and a scale mismatch

$$\Delta_p^{\text{scale}} = \left| \log_{10} \frac{\|g_p^{\text{local}}\|_2}{\|g_p^{\text{bp}}\|_2 + \varepsilon} \right|. \quad (35)$$

269 We then aggregate by parameter count across component groups (excitatory synapses, inhibitory  
 270 synapses, dendritic conductances, reactivation).

271 **Per-soma vs scalar (trained-checkpoint comparison).** Using the *trained* best Phase-2b check-  
 272 points (config.74 for MNIST, config.153 for context gating), we rerun the same one-batch

Dataset	Core	Weighted cosine	Scale mismatch	Rel. L2
MNIST	Shunting	<b>0.202</b>	<b>0.117</b>	<b>1.130</b>
MNIST	Additive	0.006	1.053	13.324
Context gating	Shunting	<b>0.108</b>	<b>0.036</b>	<b>1.404</b>
Context gating	Additive	-0.007	2.154	145.965

Table 5: **Gradient-fidelity summary (5F + per-soma broadcast).** Local vs backprop gradients compared on matched weights and batches. “Weighted cosine” is parameter-count weighted over component groups. “Scale mismatch” is  $|\log_{10}(\|g^{\text{local}}\|/\|g^{\text{bp}}\|)|$  (lower is better). Shunting networks show  $30\times$  better directional alignment and  $10\times$  lower scale distortion than additive controls.

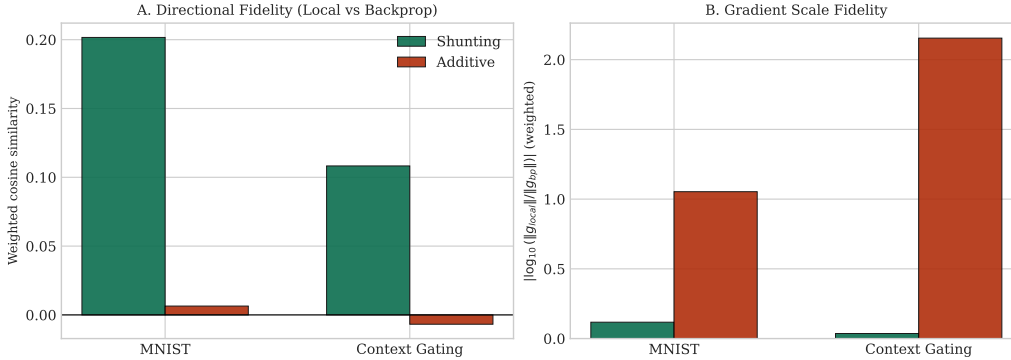


Figure 5: **Phase-2b best-regime gradient fidelity.** Shunting models show substantially better local-vs-backprop directional alignment (higher weighted cosine) and much lower gradient-scale distortion than additive controls on both MNIST and context-gating best configurations.

273 gradient-fidelity diagnostic with only `error_broadcast_mode` changed between `per_soma` and  
 274 `scalar`. For shunting cores, directional and scale metrics are nearly identical in this one-step diag-  
 275 nostic (Table 6), suggesting that the large end-task performance gap between broadcast modes is  
 276 primarily a *training-trajectory* effect rather than a single-step local-gradient direction effect.

277 **Gradient fidelity over training.** Beyond the single-step diagnostic, we track gradient alignment  
 278 throughout training by saving checkpoints every 2 epochs and computing the fidelity metrics at  
 279 each checkpoint (Fig. 7). Shunting cores maintain positive cosine similarity ( $\approx 0.05\text{--}0.15$ ) across  
 280 all epochs for the 5F rule, while additive cores fluctuate around zero. The per-component analysis  
 281 (Panel B) reveals that reactivation parameters achieve near-perfect alignment ( $> 0.99$ ) even in additive  
 282 cores, while excitatory and inhibitory synapses show the largest shunting-vs-additive gap. Scale  
 283 mismatch (Panel C) decreases during training for shunting cores, indicating that local and backprop  
 284 gradient magnitudes converge as the network learns—a form of implicit gradient normalization  
 285 enabled by the conductance-based architecture.

## 286 7.6 Core Empirical Figures

### 287 7.7 Primary Metrics and Statistical Protocol

- 288 • **Primary endpoint:** test accuracy (mean  $\pm$  95% CI over seeds).
- 289 • **Secondary:** validation NLL, convergence speed (epochs to best checkpoint), robustness  
 290 across MNIST/CIFAR-10.
- 291 • **Mechanistic:** MI/CMI terms, branch/path statistics, compartment SNR, ablation sensitivity.
- 292 • **Statistics:** two-way ANOVA (broadcast  $\times$  path), interaction contrasts, and paired seed-  
 293 matched deltas for local-vs-backprop decoder updates.

294 **Limitations and future work.** Several extensions remain: (i) scaling to deeper architectures and  
 295 image benchmarks beyond flattened inputs; (ii) extension to spiking neural networks via eligibility

Dataset (shunting)	Cosine (per-soma)	Cosine (scalar)	Scale (per-soma)	Scale (scalar)
MNIST	0.0537	0.0536	1.6756	1.6765
Context gating	0.1140	0.1138	0.8303	0.8314

Table 6: **Broadcast-mode checkpoint comparison on best Phase-2b runs.** “Scale” is  $|\log_{10}(\|g_{\text{local}}\|/\|g_{\text{bp}}\|)|$ , parameter-count weighted over component groups.

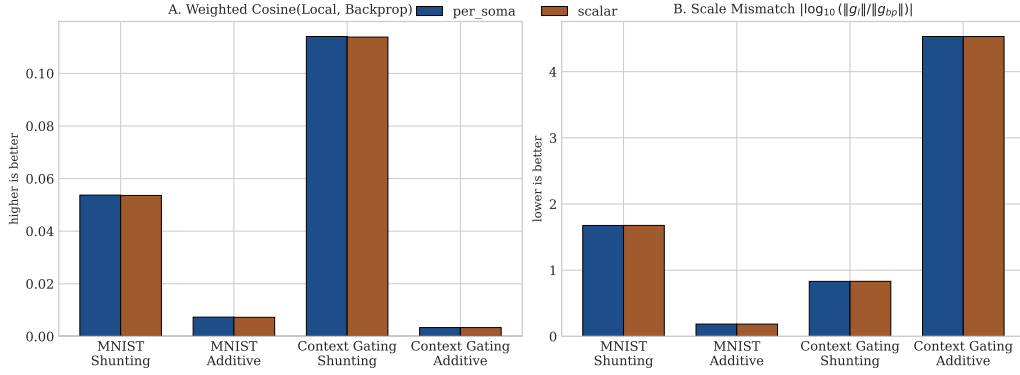


Figure 6: **Per-soma vs scalar gradient-fidelity on trained Phase-2b checkpoints.** Aggregate directional and scale metrics for shunting and additive cores on MNIST and context gating. In this one-step diagnostic, mode differences are small relative to core-type differences.

traces (Appendix C); (iii) testing on reconstructed dendritic morphologies from NeuroMorpho; and (iv) formal convergence analysis under Robbins-Monro conditions.

## 8 Related Work

**Dendritic credit assignment.** Several proposals exploit dendritic morphology for biologically plausible learning. Guerguiev et al. [7] proposed segregated dendrites where apical compartments carry error signals and basal compartments carry feedforward input, enabling approximate backpropagation in multi-layer networks. Sacramento et al. [8] showed that cortical microcircuits with dendritic predictions can approximate backprop if apical and basal inputs converge with appropriate timing. Our work differs in two key ways: (i) we derive exact gradients for conductance-based compartmental models (not abstract multi-compartment neurons), and (ii) we identify shunting inhibition as a critical enabler of local gradient quality through a quantitative gradient-fidelity analysis.

**Feedback alignment and variants.** Lillicrap et al. [5] showed that random fixed feedback weights suffice for learning in deep networks (feedback alignment, FA), and Nøkland [6] extended this to direct feedback alignment (DFA) where output errors project directly to each layer. Our broadcast error modes (Section 3.1) generalize this idea to the dendritic setting. The key difference is that our local rules also exploit conductance-based local signals (driving force, input resistance) that are unavailable to standard FA/DFA.

**Other biologically plausible methods.** Equilibrium propagation [10] computes exact gradients in energy-based networks by contrasting free and clamped phases. Predictive coding networks [9] perform inference via local prediction errors that converge to backprop gradients at equilibrium. Target propagation methods use local targets rather than error gradients. These approaches solve the weight transport problem through different mechanisms; our contribution is orthogonal, showing that dendritic biophysics provides yet another route to local credit assignment.

**Divisive normalization and shunting.** Shunting inhibition provides divisive normalization of neural responses [3], though its effect on firing rates can be subtractive in certain regimes [4]. We show that this divisive interaction has a previously unrecognized benefit for learning: it creates a regime where local gradient approximations are substantially more faithful to exact gradients.

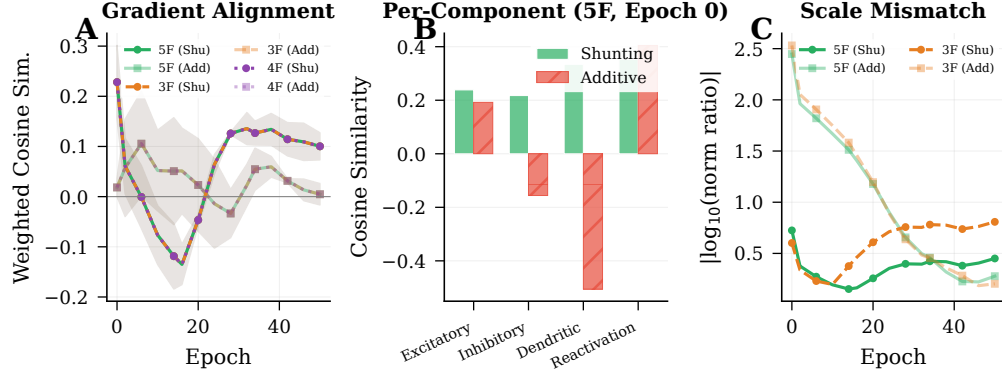


Figure 7: **Gradient fidelity over training.** (A) Parameter-count-weighted cosine similarity between local and backprop gradients over 50 training epochs. Shunting cores (solid) maintain positive alignment while additive cores (faded) fluctuate near zero. (B) Per-component cosine similarity at epoch 0 (5F rule): reactivation parameters show near-perfect alignment in both architectures, while synaptic parameters show the largest shunting advantage. (C) Scale mismatch ( $|\log_{10} \text{norm ratio}|$ ) decreases during training for shunting cores, indicating convergence of local and backprop gradient magnitudes. Shaded regions:  $\pm 1$  s.d. across 3 seeds.

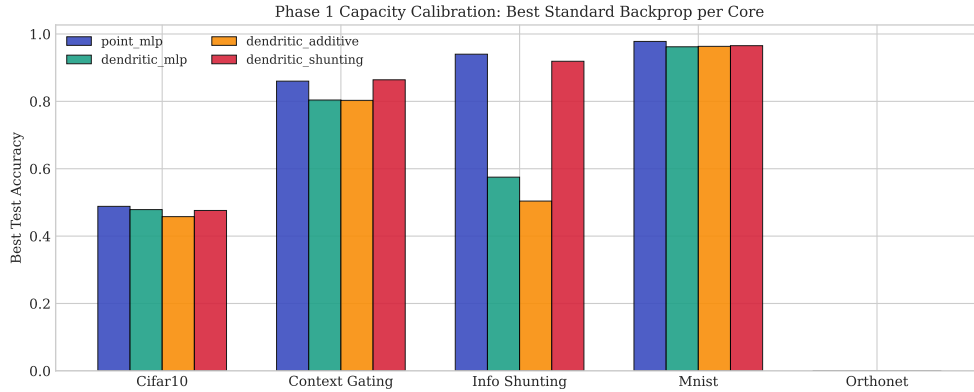


Figure 8: **Phase 1 capacity ceilings (standard backprop).** Best standard-test accuracy per dataset and core type. This establishes a meaningful performance ceiling before evaluating LocalCA competence and mechanistic effects.

323 Homeostatic mechanisms including synaptic scaling [17] and apical/basal plasticity differences [15]  
 324 motivate our morphology-aware extensions.

## 325 9 Conclusion

326 We have shown that compartmental dendritic networks with shunting inhibition create a favorable  
 327 regime for local credit assignment. Starting from conductance-based dendritic equations, we derived  
 328 exact backpropagation gradients for dendritic trees and constructed a principled hierarchy of local  
 329 approximations (3F/4F/5F) that use only synapse-local quantities plus a broadcast error signal. Our  
 330 central empirical finding is that *shunting inhibition is the key architectural enabler*: it provides  
 331 divisive normalization that dramatically improves the directional alignment and scale fidelity of local  
 332 gradients relative to exact backpropagation. The best local rule (5F with per-soma broadcast) closes  
 333 much of the gap to backpropagation on standard benchmarks, and the gradient-fidelity diagnostic  
 334 provides a new tool for understanding *why* certain architectures support local learning better than  
 335 others.

Phase 2b Gap Closing: HSIC Weight vs Error Broadcast Mode

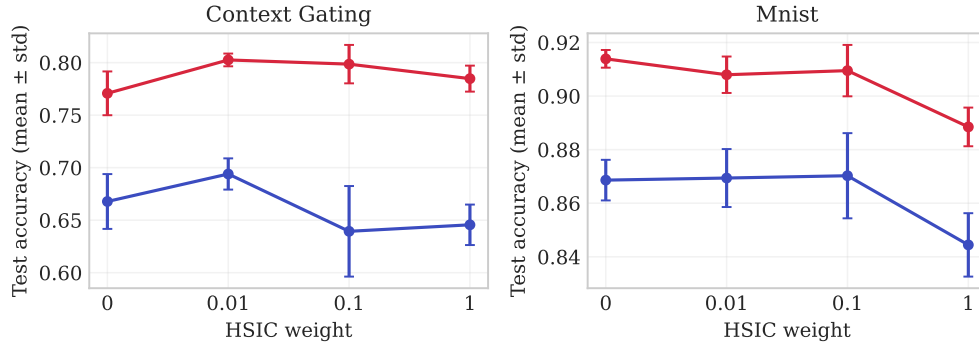


Figure 9: **HSIC strength and broadcast mode.** For context gating, moderate HSIC weights (0.01–0.1) improve local learning under per-soma broadcast, while large weights degrade performance. For MNIST, HSIC has negligible effect. Error bars:  $\pm 1$  s.d. across 5 seeds.

Decoder Locality: Local Matches Backprop, None Degrades

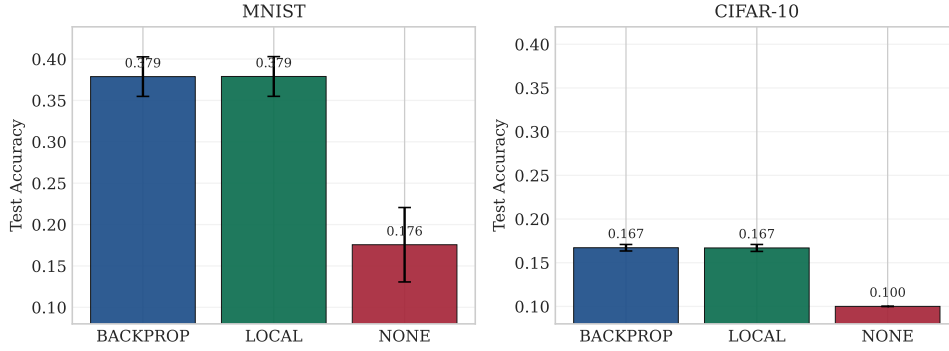


Figure 10: **Decoder locality.** On both MNIST and CIFAR-10, local decoder updates match back-propagated decoder updates, while frozen decoder weights collapse performance. This confirms that the decoder can be trained with purely local rules.

## 336 A Morphology-Aware Extensions (Details)

337 This appendix provides the full definitions for the morphology-aware extensions summarized in the  
338 main text.

339 **Branch-specific depth modulation.** Let  $d_j$  be the graph distance from the soma to branch  $j$ . Define  
340 per-branch morphology factor  $\rho_j = \rho_{\text{base}} / (d_j + \alpha)$ , where  $\alpha > 0$  prevents singularity. This mirrors  
341 cable attenuation: distal synapses receive smaller plasticity updates, with  $\|\Delta g_j^{\text{syn}}\| \propto 1 / (d_j + \alpha)$ .

342 **Dendritic normalization.** Normalize dendritic updates by total branch conductance:  $\Delta g_j^{\text{den}} \leftarrow$   
343  $\Delta g_j^{\text{den}} / (\sum_k g_k^{\text{den}} + \varepsilon)$ . This reduces update variance when total conductance  $G_n$  is large, analogous  
344 to homeostatic synaptic scaling [17].

345 **Apical vs basal branch differentiation.** Assign each branch a type flag  $t_j \in \{0, 1\}$  (basal, apical)  
346 with type-specific scales  $s_j = s_{\text{basal}} + t_j(s_{\text{apical}} - s_{\text{basal}})$ . Setting  $s_{\text{apical}} > s_{\text{basal}}$  amplifies top-  
347 down learning, consistent with distinct plasticity rules in apical vs. basal dendrites of pyramidal  
348 neurons [15].



Claim A: Shunting Advantage (Shunting - Additive)

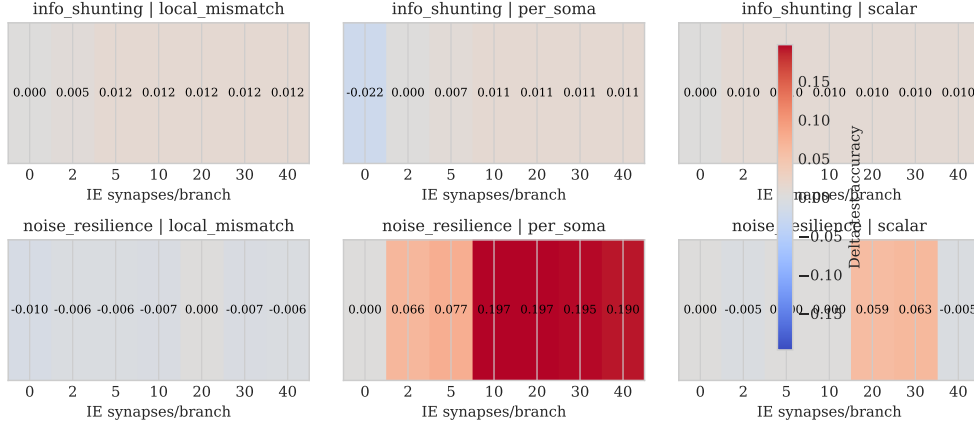


Figure 11: **Shunting advantage across inhibition strength.** Heatmap of shunting minus additive test-accuracy difference over inhibitory synapse count and broadcast modes, showing the regimes where shunting-linked local credit is most beneficial.

Regime Dependence: Shunting Dominates Additive Across Inhibition Levels

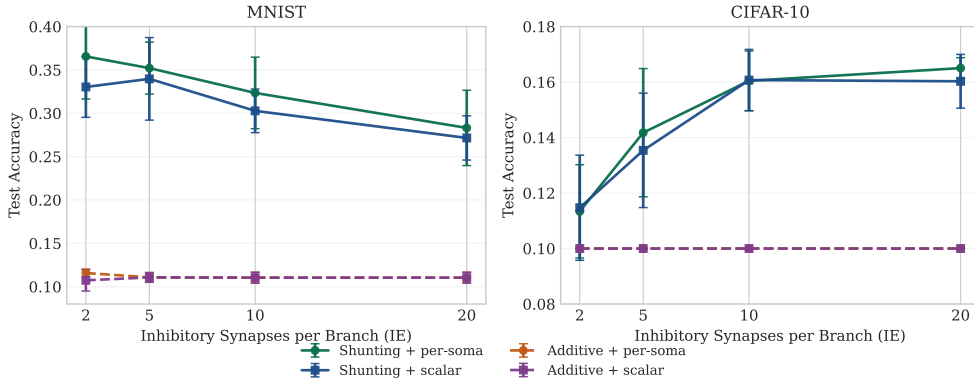


Figure 12: **Regime dependence across inhibition levels (robust sweep).** Shunting networks consistently outperform additive controls across inhibitory synapse counts and broadcast modes. This isolates the main performance source as the shunting regime rather than local-learning heuristics alone.

## 349 B HSIC Auxiliary Objectives (Details)

350 For layer activations  $\mathbf{Z} \in \mathbb{R}^{B \times d_n}$  with kernel matrix  $\mathbf{K}_Z$  and centering matrix  $\mathbf{H} = \mathbf{I} - \frac{1}{B}\mathbf{1}\mathbf{1}^\top$ , the  
 351 HSIC losses are:

$$\mathcal{L}_{\text{HSIC}}^{\text{self}} = \frac{1}{B^2} \text{tr}(\mathbf{K}_Z \mathbf{H} \mathbf{K}_Z \mathbf{H}) \quad (\text{self-decorrelation}) \quad (36)$$

$$\mathcal{L}_{\text{HSIC}}^{\text{target}} = -\frac{1}{B^2} \text{tr}(\mathbf{K}_Z \mathbf{H} \mathbf{K}_Y \mathbf{H}) \quad (\text{target-correlation}) \quad (37)$$

352 For linear kernel  $\mathbf{K}_Z = \mathbf{Z}\mathbf{Z}^\top$ , the gradients are  $\partial \mathcal{L}^{\text{self}} / \partial \mathbf{Z} = \frac{4}{B^2} \mathbf{H} \mathbf{K}_Z \mathbf{H} \mathbf{Z}$  and  $\partial \mathcal{L}^{\text{target}} / \partial \mathbf{Z} =$   
 353  $-\frac{4}{B^2} \mathbf{H} \mathbf{K}_Y \mathbf{H} \mathbf{Z}$ . These are added to synaptic eligibility traces via the chain rule.

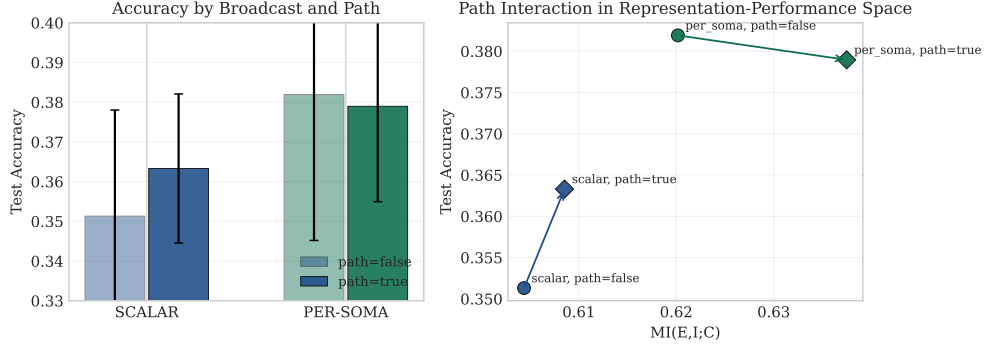


Figure 13: **Broadcast-path interaction in source analysis (robust sweep).** Path propagation has interaction-dependent behavior: within per-soma broadcast it leaves accuracy nearly unchanged while increasing information metrics, whereas within scalar broadcast it improves accuracy more with smaller information gains.

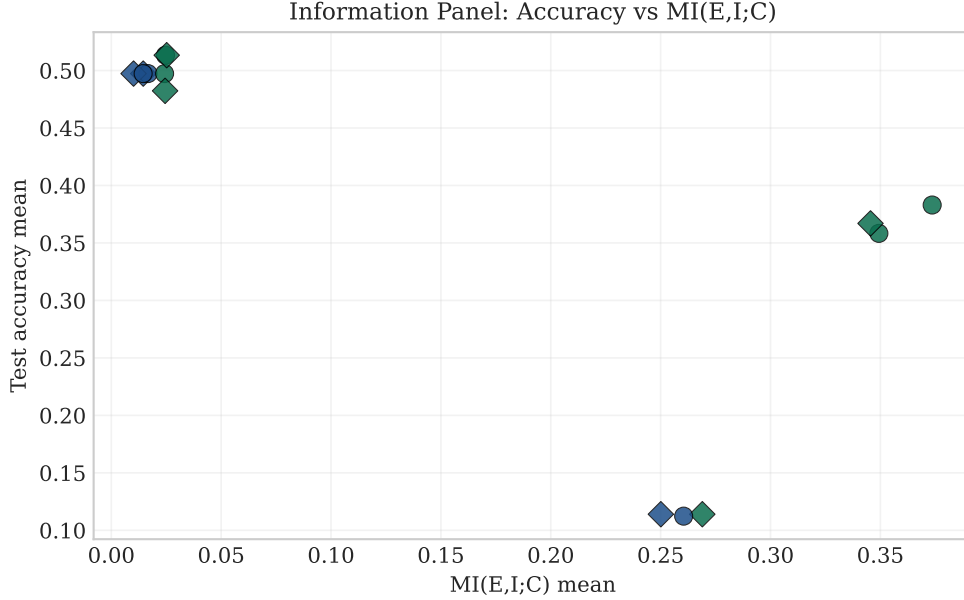


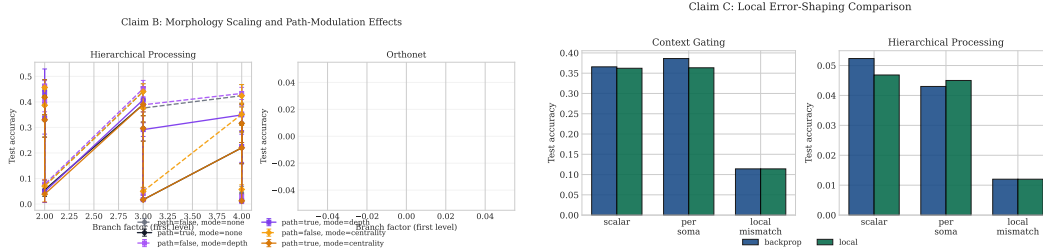
Figure 14: **Information panel.** Mutual information  $I(E, I; C)$  versus test accuracy for shunting (green) and additive (blue) networks, with (diamonds) and without (circles) path propagation. Dendritic mechanisms can systematically reshape representations even when end-task accuracy is only weakly affected, motivating mechanistic analyses beyond headline accuracy.

## 354 C Online Variant with Eligibility Traces

355 Define continuous-time eligibilities per synapse:  $\tau_e \dot{e}_j^{\text{syn}}(t) = -e_j^{\text{syn}}(t) + x_j(t) (E_j - V_n(t)) R_n^{\text{tot}}(t)$ ,  
 356 and likewise for dendritic connections. With modulatory signal  $m_n(t)$ :  $\Delta g_j^{\text{syn}} \propto \int e_j^{\text{syn}}(t) m_n(t) dt$ ,  
 357 which instantiates three-factor learning in continuous time [13, 14].

## 358 References

359 [1] Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. Oxford  
 360 University Press.



(a) **Morphology scaling.** Accuracy as a function of dendritic branching with and without path propagation/modulators. (b) **Error shaping.** Comparison of broadcast modes and decoder update modes on hierarchical/context tasks.

Figure 15: **Phase 3 ablations: morphology and error shaping.** These sweeps identify which LocalCA components and morphology-aware extensions have the largest effect on performance and representation metrics.

Robust Claim / Condition	Metric	Value
Claim2 (MNIST, decoder local vs backprop vs none)	test accuracy	0.3790 vs 0.3788 vs 0.1756
Claim2 (CIFAR-10, decoder local vs backprop vs none)	test accuracy	0.1669 vs 0.1671 vs 0.1000
Claim3 (MNIST, shunting vs additive; avg matched)	test accuracy delta	+0.210
Claim3 (CIFAR-10, shunting vs additive; avg matched)	test accuracy delta	+0.044
Claim4 (per-soma, path true minus false)	test / $MI(E, I; C)$	-0.0030 / +0.0173

Table 7: **Mechanistic ablation results (controlled architecture).** These experiments use a smaller architecture to cleanly isolate individual mechanisms (decoder locality, shunting regime dependence, broadcast/path interaction) rather than maximize absolute performance.

- [2] Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. MIT Press.
- [3] Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1), 51–62.
- [4] Holt, G. R., & Koch, C. (1997). Shunting inhibition does not have a divisive effect on firing rates. *Neural Computation*, 9(5), 1001–1013.
- [5] Lillicrap, T. P., Cownden, D., Tweed, D. B., & Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7, 13276.
- [6] Nøkland, A. (2016). Direct feedback alignment provides learning in deep neural networks. *Advances in Neural Information Processing Systems*, 29.
- [7] Guerguiev, J., Lillicrap, T. P., & Richards, B. A. (2017). Towards deep learning with segregated dendrites. *eLife*, 6, e22901.
- [8] Sacramento, J., Costa, R. P., Bengio, Y., & Senn, W. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. *Advances in Neural Information Processing Systems*, 31.
- [9] Whittington, J. C., & Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends in Cognitive Sciences*, 23(3), 235–250.
- [10] Scellier, B., & Bengio, Y. (2017). Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in Computational Neuroscience*, 11, 24.
- [11] Gretton, A., Bousquet, O., Smola, A., & Schölkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. *International Conference on Algorithmic Learning Theory*, 63–77.
- [12] Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., & Smola, A. J. (2007). A kernel statistical test of independence. *Advances in Neural Information Processing Systems*, 20.

- 384 [13] Frémaux, N., & Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and  
385 theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9, 85.
- 386 [14] Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., & Maass, W.  
387 (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature*  
388 *Communications*, 11, 3625.
- 389 [15] Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for  
390 the cerebral cortex. *Trends in Neurosciences*, 36(3), 141–151.
- 391 [16] Welford, B. P. (1962). Note on a method for calculating corrected sums of squares and products.  
392 *Technometrics*, 4(3), 419–420.
- 393 [17] Turrigiano, G. G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*,  
394 135(3), 422–435.