

---

# Local Credit Assignment in Compartmental Dendritic Networks

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

How can neurons learn efficiently when plasticity is synapse-local and global supervision is low bandwidth? We study conductance-based compartmental dendritic networks and show that dendritic structure and shunting inhibition create regimes where strictly local learning becomes effective and mechanistically interpretable. Starting from the compartment voltage equation, we derive exact loss gradients for arbitrary dendritic trees and obtain a factorization that highlights synapse-local terms (presynaptic drive, driving force, input resistance) and a global modulatory term (broadcast error). This motivates a hierarchy of local rules—3-factor (3F), morphology-modulated 4-factor (4F), and information-modulated 5-factor (5F) updates—plus morphology-aware and shunting-aware extensions. Empirically, shunting inhibition is the key architectural enabler: across completed sweeps it yields large gains over additive controls and markedly improves local-vs-backprop gradient fidelity (direction and scale). Under a calibrated capacity regime, the best local configuration (5F with per-soma broadcast) reaches  $0.914 \pm 0.003$  test accuracy on MNIST and  $0.803 \pm 0.006$  on context gating, compared to backprop ceilings of 0.965 and 0.864 on the same shunting architecture.

## 1 Introduction

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

This paper asks a concrete question: *can dendritic structure and shunting inhibition create regimes where strictly local learning rules (synapse-local factors + a broadcast teaching signal) approach the credit-assignment quality of backpropagation?* We answer affirmatively. Starting from conductance-based dendritic equations, we derive exact gradients for dendritic trees and construct a hierarchy of biologically-local approximations (3-factor, 4-factor, and 5-factor rules). We then show empirically that *shunting inhibition*—conductance that primarily modulates input resistance and gain—is the key architectural enabler for local learning: it stabilizes credit signals and substantially improves local-vs-backprop gradient fidelity compared to additive dendritic controls.

## Contributions.

- 34 1. **Exact gradients for compartmental dendritic trees.** We derive the exact loss gradient for  
 35 arbitrary dendritic tree morphologies in a conductance-based compartment model, making  
 36 explicit the multiplicative path factors that standard backprop implicitly computes.
- 37 2. **A unified local-rule template (3F/4F/5F + modifiers).** We express a family of strictly  
 38 local updates in a shared factorized form, separating synapse-local terms (presynaptic drive,  
 39 driving force, input resistance) from global teaching terms (broadcast errors) and optional  
 40 morphology/information modulators.
- 41 3. **Shunting as an architectural enabler of local learning.** We show that shunting inhibition  
 42 yields large and regime-dependent benefits for local learning and that these gains are  
 43 accompanied by substantially improved local-vs-backprop gradient fidelity compared to  
 44 additive dendritic controls.
- 45 4. **Mechanistic diagnostics beyond accuracy.** We introduce a component-wise gradient-  
 46 fidelity diagnostic (direction and scale) that links architecture and learning-rule design to  
 47 credit-signal quality.

48 **Positioning.** Our approach occupies a unique position in the landscape of biologically plausi-  
 49 ble learning (Table 3). While most dendritic credit assignment models use abstract compartmen-  
 50 tal surrogates—segregated dendrites [11], burst-based signaling [22], or microcircuit prediction  
 51 errors [12]—we derive learning rules directly from conductance-based voltage equations where  
 52 shunting inhibition arises naturally as divisive normalization. This bridges two literatures that have  
 53 developed largely in parallel: (i) canonical divisive normalization in sensory processing [6], and  
 54 (ii) biologically plausible credit assignment, where the role of inhibitory conductances in gradient  
 55 quality has not been explored. Our gradient-fidelity diagnostic (Section 5.3) provides the mechanistic  
 56 link, demonstrating quantitatively that shunting architecture improves the direction and scale of local  
 57 credit signals relative to additive controls.

## 58 **Three empirical findings (this draft).**

- 59 1. **Local competence:** In a calibrated capacity regime, 5F with per-soma broadcast closes  
 60 much of the backprop gap on both MNIST and context gating (Section 5).
- 61 2. **Regime dependence:** Shunting dendritic cores outperform additive controls under the same  
 62 local rule, with gaps widening under stronger inhibition/noise (Fig. 2).
- 63 3. **Mechanism:** Shunting architectures yield substantially higher local-vs-backprop gradient  
 64 fidelity (Table 2, Fig. 3), supporting a mechanistic explanation for the performance gains.

## 65 **2 Compartmental Voltage Model**

66 We use a standard steady-state conductance model obtained by discretizing passive cable dynamics  
 67 (e.g., [1, 2]). In normalized units with leak reversal potential 0 and unit leak conductance, each  
 68 compartment voltage is a conductance-weighted average of synaptic reversal potentials, child voltages,  
 69 and leak. This form makes two facts explicit: (i) local sensitivities depend on the driving force  
 70 ( $E - V$ ) and input resistance  $R^{\text{tot}}$ , and (ii) shunting inhibition corresponds to adding conductance  
 71 with  $E_{\text{inh}} \approx 0$  (Section 2.3).

### 72 **2.1 Voltage Equation**

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

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

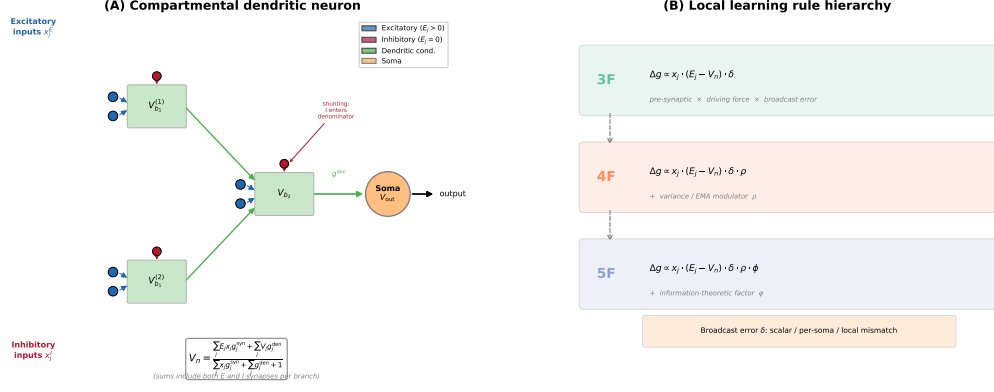


Figure 1: **Model overview.** (A) A compartmental dendritic neuron where *each branch* receives both excitatory ( $E_j > 0$ , blue) and inhibitory ( $E_j = 0$ , red) synaptic inputs via separate sparse connectivity (TopK). Inhibitory conductances enter only the denominator of the voltage equation (shunting/divisive 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.

80 **Currents.** Synaptic current:

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

81 Dendritic current:

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

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

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

**Total conductance and resistance.**

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

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

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

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

## 86 2.2 Local Sensitivities

**Proposition 1** (Synaptic Gradient).

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

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

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

**Proposition 3** (Dendritic Gradient).

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

## 87 2.3 Shunting Inhibition and Divisive Gain Control

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

90 **Proposition 4** (Subthreshold Effect of Shunts). *For a pure shunt ( $E_j = 0$ ), the steady-state sensitivity*  
91 *to the inhibitory conductance is*

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

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

94 **Remark 1** (Divisive vs. Subtractive at the Firing-Rate Level). *While shunting produces divisive*  
95 *scaling of subthreshold voltages, its net effect on firing rates can be subtractive in many regimes*  
96 *[7], so we report both voltage- and rate-level analyses in experiments. Normalization via added*  
97 *conductance is consistent with canonical divisive normalization models in cortex [6].*

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

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

99 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  
100 that the same multiplicative factors are applied to both excitatory and inhibitory synapses in the  
101 implementation; the sign difference arises solely from the driving force ( $E_j - V_n$ ).

## 102 2.4 Loss Propagation

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

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

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

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

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

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

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

## 112 3 Local Learning Approximations

### 113 3.1 Broadcast Error Approximation

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

$$\frac{\partial L}{\partial V_n} \approx e_n, \quad \prod_{i=0}^n R_i^{\text{tot}} g_i^{\text{den}} \approx 1 \quad (11)$$

116 Three broadcast modes are considered:

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

$$\bar{\delta}(b) = \frac{1}{d_{\text{out}}} \sum_{k=1}^{d_{\text{out}}} \delta_k(b), \quad e_n(b) = \bar{\delta}(b) \mathbf{1}_{d_n} \quad (12)$$

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

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

$$\varepsilon_n(b) = (P_n(b) - V_n(b)) - \frac{1}{B} \sum_{t=1}^B (P_n(t) - V_n(t)) \quad (13)$$

123 Then:

$$e_n(b) = \bar{\delta}(b) \varepsilon_n(b) \quad (14)$$

### 124 3.2 Gradient Alignment with Broadcast Errors

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

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

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

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

133 *Sketch.* Adapt the feedback-alignment argument [9, 10]: fixed random feedback suffices for align-  
 134 ment as forward weights adapt. Here,  $\hat{\Xi}_n$  is proportional to  $\Xi_n$  up to the missing path factor; Jensen  
 135 bounds on (10) yield  $c_n > 0$ .  $\square$

### 136 3.3 Three-Factor Rule (3F)

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

$$\Delta g_j^{\text{syn}} = \eta \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B \quad (15)$$

138 *For dendritic conductances:*

$$\Delta g_j^{\text{den}} = \eta \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B \quad (16)$$

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

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

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

### 144 3.4 Four-Factor Rule (4F): Morphology Correlation

145 **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*  
 146 *in layer  $n$ . Define the correlation with output:*

$$\rho_n = \frac{\text{Cov}(\bar{V}_n, \bar{V}_0)}{\sqrt{\text{Var}(\bar{V}_n) \text{Var}(\bar{V}_0) + \varepsilon}} \quad (17)$$

147 **Proposition 5** (4F Update Rule). *Multiply 3F updates by  $\rho_n$ :*

$$\Delta g_j^{\text{syn}} = \eta \rho_n \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B \quad (18)$$

$$\Delta g_j^{\text{den}} = \eta \rho_n \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B \quad (19)$$

148 **Proposition 6** (Approximate Gradient Alignment). *Let  $L$  be a smooth loss. If layer  $n$  contributes to*  
 149 *the output primarily through its mean activity, then  $\mathbb{E}[\frac{\partial L}{\partial V_n} \cdot \bar{V}_n] \propto \rho_n \cdot \text{Var}(\bar{V}_n)$ . Thus  $\rho_n$  weights*  
 150 *updates by the layer’s relevance to the task.*

151 **Estimators (EMA / online).** For minibatches  $B \geq 2$ , estimate  $\rho_n$  from (17) with an EMA over  
 152 batches. For  $B = 1$  (online), maintain means  $\mu_x, \mu_y$ , variances  $\sigma_x^2, \sigma_y^2$ , and covariance  $C_{xy}$  for  
 153  $x_t = \bar{V}_0^{(t)}$  and  $y_t = \bar{V}_n^{(t)}$  using Welford’s numerically stable algorithm [20]:

$$\begin{aligned} \mu_x^{(t)} &= (1 - \alpha) \mu_x^{(t-1)} + \alpha x_t, & \mu_y^{(t)} &= (1 - \alpha) \mu_y^{(t-1)} + \alpha y_t, \\ \delta_x &= x_t - \mu_x^{(t-1)}, & \delta_y &= y_t - \mu_y^{(t-1)}, \\ \sigma_x^{2(t)} &= (1 - \alpha) \sigma_x^{2(t-1)} + \alpha \delta_x^2, & \sigma_y^{2(t)} &= (1 - \alpha) \sigma_y^{2(t-1)} + \alpha \delta_y^2, \\ C_{xy}^{(t)} &= (1 - \alpha) C_{xy}^{(t-1)} + \alpha \delta_x \delta_y. \end{aligned} \quad (20)$$

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

### 155 3.5 Five-Factor Rule (5F): Conditional Information

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

$$\beta_n = \frac{\text{Cov}(V_n, P_n)}{\text{Var}(P_n) + \lambda} \quad (21)$$

$$\sigma_{\text{res}}^2 = \text{Var}(V_n) - \beta_n \text{Cov}(V_n, P_n) \quad (22)$$

158 *The information proxy is:*

$$\phi_n = \frac{\text{Var}(V_n)}{\sigma_{\text{res}}^2 + \varepsilon} = \frac{1}{1 - R_n^2} \geq 1, \quad (23)$$

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

160 **Remark 3** (Information-Theoretic Interpretation).  $\phi_n$  increases when  $V_n$  is more predictable from  
 161 its parent  $P_n$  (higher  $R^2$ ), amplifying updates for compartments with strong signal propagation.  
 162 In practice,  $\phi_n$  is clamped to  $[0.25, 4.0]$  for stability. An alternative  $\phi_n = 1 - R^2$  would instead  
 163 emphasize compartments with unique information; both are valid depending on whether coherent  
 164 signal flow or novelty is prioritized.

**Proposition 7** (5F Update Rule).

$$\Delta g_j^{\text{syn}} = \eta \rho_n \phi_n \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B \quad (24)$$

$$\Delta g_j^{\text{den}} = \eta \rho_n \phi_n \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B \quad (25)$$

## 165 4 Morphology-Aware Extensions

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

### 168 4.1 Path-Integrated Propagation

169 Exact tree backpropagation contains a path-sum of multiplicative edge factors (Eq. 10), which induces  
 170 depth-dependent attenuation. We approximate this attenuation with a per-layer *path factor*  $\pi_n$  and  
 171 modulate the broadcast error as  $e_n \leftarrow e_n \cdot \pi_n$ :

$$\pi_n = \begin{cases} 1 & n = 0 \\ \pi_{n-1} \cdot R_{n-1}^{\text{tot}} \cdot \bar{g}_{n-1}^{\text{den}} & n \geq 1, \end{cases} \quad (26)$$

172 where  $\bar{g}_{n-1}^{\text{den}}$  is the mean dendritic conductance from layer  $n - 1$  to  $n$ . In practice,  $\pi_n$  is computed per  
 173 sample and broadcast within each layer for stability (Appendix D).

Dataset	Backprop ceiling (shunting)	Best LocalCA (5F, per-soma)	Gap
MNIST	0.965	$0.914 \pm 0.003$	-0.051
Context gating	0.864	$0.803 \pm 0.006$	-0.061

Table 1: **Local competence in a calibrated capacity regime.** Backprop ceilings are computed from Phase-1 capacity sweeps; LocalCA values are the best completed Phase-2b results within the same architecture regime. Errors are across seeds.

## 4.2 Additional Extensions

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

- **Depth modulation:** Per-branch scaling  $\rho_j = \rho_{\text{base}}/(d_j + \alpha)$  that mirrors cable attenuation, biasing learning toward proximal synapses.
- **Dendritic normalization:** Update normalization by total branch conductance,  $\Delta g_j^{\text{den}} \leftarrow \Delta g_j^{\text{den}} / (\sum_k g_k^{\text{den}} + \varepsilon)$ , which stabilizes update variance analogous to homeostatic scaling [21].
- **Apical/basal differentiation:** Branch-type-specific scaling factors  $s_j$  that allow differential plasticity in feedback (apical) vs. feedforward (basal) compartments [19].

**HSIC auxiliary objectives.** We optionally add Hilbert-Schmidt Independence Criterion (HSIC) losses [15] as auxiliary objectives: a self-decorrelation term that encourages diverse representations within each layer, and a target-correlation term that aligns layer activations with labels. Moderate HSIC weights improve performance on context-gating tasks while having negligible effect on MNIST (see Section 5). Full definitions and gradients are in Appendix E.

## 5 Experiments

### 5.1 Setup

We evaluate local credit assignment in two complementary regimes: (i) a *capacity-calibrated* regime where standard backprop achieves high accuracy on the same architectures used for local learning, and (ii) a *controlled small-network* sandbox used to isolate mechanisms (decoder locality, inhibition sweeps, and broadcast/path interactions). We report MNIST and context gating as primary datasets, with CIFAR-10 (flattened) used only as a sanity check for the decoder-locality and shunting-regime claim sweeps (Appendix A). Architectures include point MLP baselines and dendritic cores with either additive dendritic integration or shunting (conductance-based) inhibition. For each setting, we compare standard backprop training to LocalCA training under matched optimization protocols.

### 5.2 Results: Three Findings

**Finding 1: Local competence under calibrated capacity.** In a Phase-1 capacity calibration sweep, standard training achieves high ceilings on MNIST and context gating with dendritic shunting cores (0.965 and 0.864 test, respectively; Appendix A). Within this same capacity regime, Phase-2b local-competence sweeps show that the 5F family is consistently strongest, and that per-soma broadcast is a critical factor (Appendix A). Table 1 summarizes the resulting gap closing for the best completed LocalCA configurations.

**Finding 2: Regime dependence across inhibition strength.** Local learning gains are not uniform: they concentrate in inhibition/noise-stress regimes where shunting inhibition is active. Figure 2 shows the shunting-minus-additive advantage over inhibitory synapse count and broadcast mode; the gap widens as inhibition increases, consistent with shunting-linked divisive gain control stabilizing credit signals. Error-shaping ablations indicate that this architectural advantage depends strongly on broadcast design: per-soma remains consistently strong, whereas local-mismatch remains substantially weaker even after a post-fix recheck. We report full numerical details in Appendix A, Table 5, and Fig. 14.

## Claim A: Shunting Advantage Across Inhibitory Regimes

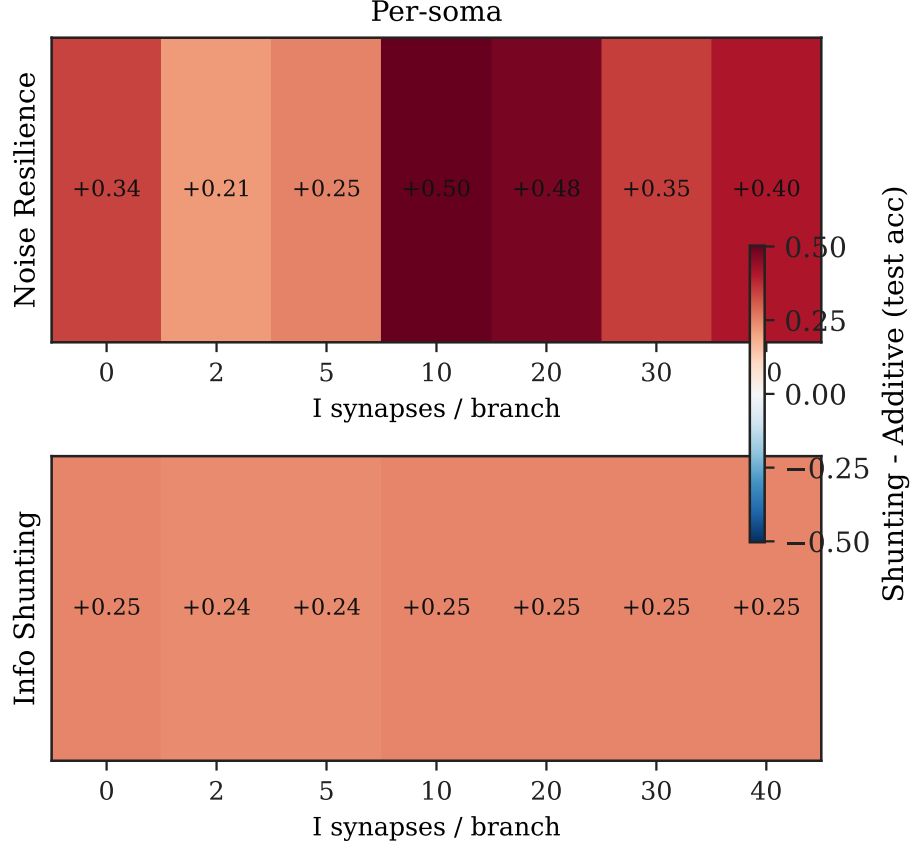


Figure 2: **Shunting advantage across inhibition strength.** Heatmap of shunting minus additive test-accuracy difference over inhibitory synapse count and broadcast modes, showing the regimes where shunting-linked local credit is most beneficial.

**Finding 3: Shunting improves credit signal fidelity.** To test whether performance gains correspond to higher-quality credit signals, we compare LocalCA and backprop gradients on the *same batch and same weights* using a component-wise gradient-fidelity diagnostic (Section 5.3). Shunting architectures show substantially higher directional alignment and lower scale mismatch than additive controls on both MNIST and context gating (Table 2, Fig. 3).

### 5.3 Gradient-Fidelity Diagnostic (Local vs Backprop)

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

$$\cos_p = \frac{\langle g_p^{\text{local}}, g_p^{\text{bp}} \rangle}{\|g_p^{\text{local}}\|_2 \|g_p^{\text{bp}}\|_2 + \varepsilon}, \quad (27)$$

and a scale mismatch

$$\Delta_p^{\text{scale}} = \left| \log_{10} \frac{\|g_p^{\text{local}}\|_2}{\|g_p^{\text{bp}}\|_2 + \varepsilon} \right|. \quad (28)$$

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



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 2: **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.

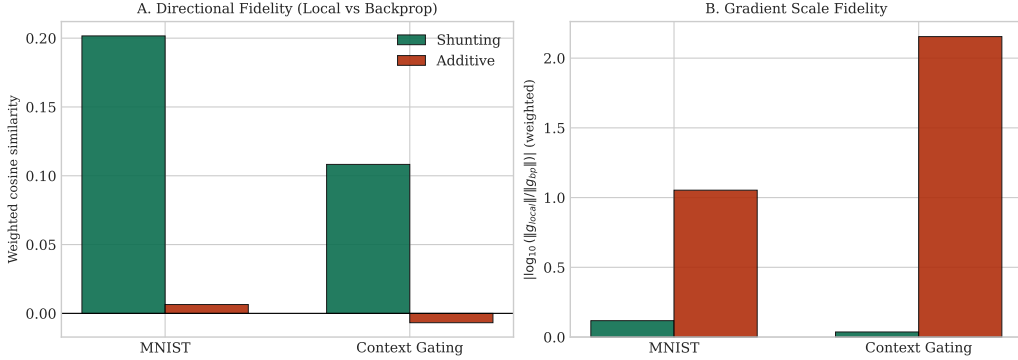


Figure 3: **Phase-2b best-regime gradient fidelity.** Shunting models show substantially better local-vs-backprop directional alignment (higher weighted cosine) and much lower gradient-scale distortion than additive controls on both MNIST and context-gating best configurations.

226 **Gradient alignment dynamics over training.** Beyond the static snapshot, we track how local-vs-backprop alignment evolves during training. Figure 4 shows per-layer weighted cosine similarity over epochs. In shunting networks, alignment at the proximal (soma-adjacent) layer approaches  $\sim 1.0$  and improves steadily during training, while distal layers show modest positive alignment. Additive networks, by contrast, show near-zero or slightly negative alignment across all layers and epochs. Figure 5 decomposes this spatially: after training, shunting networks show increasing alignment from distal to proximal layers (reaching near-perfect at layer 2), whereas additive networks show uniformly poor alignment. The right panel reveals the underlying mechanism: additive networks exhibit  $> 3\times$  larger scale mismatch at initialization, which improves only at distal layers during training; shunting networks maintain low scale mismatch throughout.

236 **Component-wise alignment dynamics.** Figure 6 decomposes gradient alignment by parameter type. In shunting networks, dendritic conductances and E-synapses show the highest alignment ( $\sim 0.3$ – $0.5$ ), consistent with the biophysical role of conductance-based driving forces in shaping local credit signals. Reactivation parameters show more variable alignment, while I-synapses contribute less. In additive networks, no component achieves sustained positive alignment, confirming that the credit-signal advantage of shunting is not confined to a single parameter class.

242 **Limitations and future work.** We view these results as evidence for a mechanism (shunting-linked credit stabilization) rather than a complete biological account. A key limitation is that local-mismatch broadcast, although partially stabilized by our implementation update, still underperforms substantially relative to per-soma broadcast in the tested regimes (Appendix A, Table 5). Thus, our strongest claims are about 5F + per-soma in conductance-based shunting networks, not about all local broadcast constructions. Remaining extensions include scaling to deeper architectures and vision benchmarks beyond flattened inputs, testing on reconstructed morphologies, redesigning mismatch-style local broadcasts, and connecting the discrete-time rules to event-driven spiking implementations (Appendix F).

## Gradient alignment dynamics over training

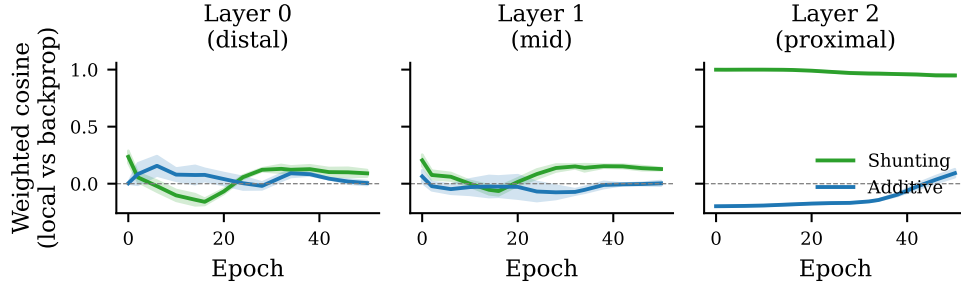


Figure 4: **Gradient alignment dynamics over training.** Per-layer weighted cosine similarity between local and backprop gradients across epochs. Shunting (green) shows strong alignment at the proximal layer and gradual improvement at deeper layers; additive (blue) remains near zero. Shaded regions:  $\pm 1$  s.e. across seeds.

## Layer-by-layer gradient quality

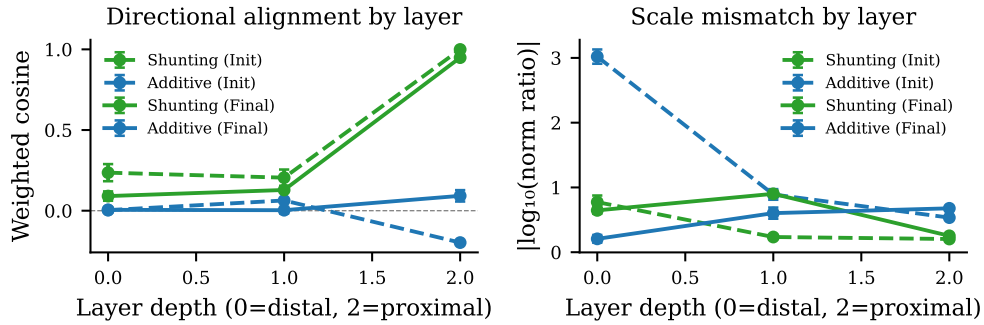


Figure 5: **Layer-by-layer gradient quality.** Left: directional alignment (weighted cosine) increases from distal to proximal layers in shunting networks (green), especially after training (solid). Right: gradient scale mismatch ( $|\log_{10} \text{norm ratio}|$ ) is uniformly low for shunting but high for additive at distal layers. Dashed: initial weights; solid: final weights.

## 6 Related Work

**Dendritic models of credit assignment.** Dendritic trees support nonlinear subunit computation [3, 4] and have inspired several biologically plausible learning schemes. Urbanczik & Senn [5] derive a single-neuron rule from dendritic prediction of somatic spiking. Guerguiev et al. [11] propose segregated dendrites where apical compartments carry teaching signals and basal compartments carry feedforward input. Sacramento et al. [12] show that cortical microcircuits with dendritic prediction errors can approximate backpropagation (98.0% MNIST). Payeur et al. [22] introduce burst-dependent plasticity where apical burst rates carry top-down error signals, and Greedy et al. [23] extend this in BurstCCN with cortical microcircuit constraints. Haider et al. [24] combine dendritic compartments with prospective coding in Latent Equilibrium (98.9% MNIST). Richards & Lillicrap [25] provide a comprehensive review of dendritic solutions to credit assignment. Our work differs from all the above in two key respects: (i) we derive exact gradients from a *conductance-based* compartment model (Eq. 3–10) rather than abstract compartmental surrogates, and (ii) we identify shunting inhibition as a critical enabler of local gradient quality via a quantitative gradient-fidelity diagnostic—a connection between normalization and learning that has not been previously established.

**Feedback alignment and local learning methods.** Lillicrap et al. [9] showed that random fixed feedback weights suffice for learning (feedback alignment, FA), and Nøkland [10] extended this to

## Component-wise gradient alignment dynamics

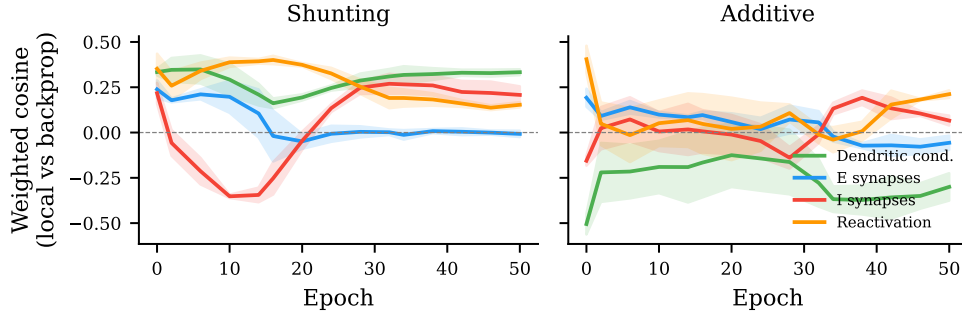


Figure 6: **Component-wise gradient alignment dynamics.** Per-component weighted cosine (local vs backprop) over epochs for shunting (left) and additive (right) networks. Dendritic conductances (green) and E-synapses (blue) carry the strongest alignment signal in shunting networks.

268 direct feedback alignment (DFA) where output errors project directly to each layer. The forward-  
 269 forward algorithm [26] replaces backpropagation entirely with layer-local contrastive objectives.  
 270 PEPITA [27] achieves learning through forward-only error propagation. These methods demonstrate  
 271 that local or random-feedback learning is feasible in standard architectures (typically 97–99% on  
 272 MNIST MLPs). Our broadcast error modes (Section 3.1) generalize FA/DFA to the dendritic setting,  
 273 but our local rules additionally exploit conductance-based signals—driving force ( $E_j - V_n$ ) and input  
 274 resistance  $R_n^{\text{tot}}$ —that arise naturally from the biophysics and are unavailable to standard FA/DFA.

275 **Target propagation and energy-based methods.** Difference target propagation [28] replaces  
 276 error gradients with layer-wise target activations. Meulemans et al. [29] introduce Difference  
 277 Feedback Control (DFC) with controller-based credit assignment (97.8% MNIST, 55.3% CIFAR-10).  
 278 Equilibrium propagation [14] computes exact gradients in energy-based networks by contrasting  
 279 free and clamped phases. Predictive coding networks [13, 30] perform inference via local prediction  
 280 errors that converge to backprop gradients at equilibrium. These methods solve the weight transport  
 281 problem through diverse mechanisms. Our contribution is orthogonal: we show that conductance-  
 282 based dendritic biophysics provides an additional route to local credit, with the unique advantage of a  
 283 mechanistic gradient-fidelity diagnostic that links architecture to credit-signal quality.

284 **Divisive normalization and dendritic biophysics.** Shunting inhibition implements divisive nor-  
 285 malization of neural responses [6], though its effect on firing rates can be subtractive in certain  
 286 regimes [7]. Koch, Poggio, & Torre [31] showed that shunting interactions enable AND/OR-like  
 287 logic in dendritic trees. Silver [32] demonstrated that inhibitory conductance modulates gain, dy-  
 288 namic range, and signal-to-noise ratio across neural circuits. Beniaguev et al. [33] showed that  
 289 a biophysical neuron with active dendrites is computationally equivalent to a 5–8 layer deep neu-  
 290 ral network, underscoring single-neuron computational power. Recent in vivo work demonstrates  
 291 compartment-specific plasticity rules in cortical neurons, with distinct learning dynamics in basal  
 292 versus apical dendrites [19]. Inhibitory plasticity and E/I balance are also critical for stabilizing  
 293 cortical dynamics [8], while homeostatic mechanisms including synaptic scaling [21] motivate our  
 294 morphology-aware extensions. Despite extensive work on divisive normalization *and* on biologically  
 295 plausible learning separately, *no prior work has explicitly connected shunting inhibition to gradient*  
 296 *computation or credit assignment quality*. This is the central gap our work fills: we show that the same  
 297 conductance-based mechanism known for sensory normalization also creates favorable conditions for  
 298 local credit propagation.

299 **Broader relevance.** Our results speak to several communities. For *computational neuroscience*,  
 300 the gradient-fidelity diagnostic provides a new tool for evaluating how biophysical architecture  
 301 shapes learning. For *theoretical neuroscience*, we identify a previously unexplored function of  
 302 divisive normalization: improving the fidelity of locally computed credit signals, extending its known  
 303 roles in gain control [32] and sensory coding [6]. For *machine learning*, the conductance-based

Method	Paradigm	MNIST	Cond.	Diag.
<i>Generic local learning</i>				
FA [9]	Random feedback	97–98%		
DFA [10]	Direct feedback	97.3%		
Forward-forward [26]	Goodness	98.6%		
PEPITA [27]	Forward error	98.0%		
<i>Target / energy-based</i>				
DTP [28]	Difference targets	98.5%		
DFC [29]	Feedback control	97.8%		
EP [14]	Energy contrast	97–98%		
PC [13]	Prediction error	98.5%		
<i>Dendritic</i>				
Sacramento et al. [12]	Microcircuit	98.0%	○	
Payeur et al. [22]	Burst signaling	97.5%	○	
BurstCCN [23]	Cortical burst	~97%	○	
Latent EQ [24]	Prospective	98.9%	○	
<b>Ours (5F)</b>	<b>Conductance</b>	<b>91.4%</b>	<b>●</b>	<b>●</b>

Table 3: **Landscape of biologically plausible learning** (MLP on MNIST). **Cond.:** ● = derived from conductance-based biophysics; ○ = uses dendritic compartments as abstract surrogates. **Diag.:** provides quantitative local-vs-backprop gradient-fidelity analysis. Our method operates in conductance-based voltage space (not standard activation space), which constrains raw accuracy but enables a mechanistic link between shunting normalization and credit-signal quality that other methods do not provide. All MNIST numbers are for MLP architectures from published results.

framework suggests that architectural inductive biases inspired by dendritic biophysics can shape gradient geometry in ways that benefit local learning. For *neuromorphic engineering*, the strictly local nature of our rules (synapse-local factors + low-bandwidth broadcast) maps naturally onto parallel neuromorphic substrates where global error transport is costly.

## 7 Conclusion

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

## References

- [1] Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press.
- [2] Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. MIT Press.
- [3] Poirazi, P., Brannon, T., & Mel, B. W. (2003). Pyramidal neuron as two-layer neural network. *Neuron*, 37(6), 989–999.
- [4] London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience*, 28, 503–532.
- [5] Urbanczik, R., & Senn, W. (2014). Learning by the dendritic prediction of somatic spiking. *Neuron*, 81(3), 521–528.

- [6] Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1), 51–62.
- [7] Holt, G. R., & Koch, C. (1997). Shunting inhibition does not have a divisive effect on firing rates. *Neural Computation*, 9(5), 1001–1013.
- [8] Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., & Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334(6062), 1569–1573.
- [9] 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.
- [10] Nøkland, A. (2016). Direct feedback alignment provides learning in deep neural networks. *Advances in Neural Information Processing Systems*, 29.
- [11] Guerguiev, J., Lillicrap, T. P., & Richards, B. A. (2017). Towards deep learning with segregated dendrites. *eLife*, 6, e22901.
- [12] Sacramento, J., Costa, R. P., Bengio, Y., & Senn, W. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. *Advances in Neural Information Processing Systems*, 31.
- [13] Whittington, J. C., & Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends in Cognitive Sciences*, 23(3), 235–250.
- [14] Scellier, B., & Bengio, Y. (2017). Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in Computational Neuroscience*, 11, 24.
- [15] Gretton, A., Bousquet, O., Smola, A., & Schölkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. *International Conference on Algorithmic Learning Theory*, 63–77.
- [16] 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.
- [17] Frémaux, N., & Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9, 85.
- [18] Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., & Maass, W. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11, 3625.
- [19] Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in Neurosciences*, 36(3), 141–151.
- [20] Welford, B. P. (1962). Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4(3), 419–420.
- [21] Turrigiano, G. G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*, 135(3), 422–435.
- [22] Payeur, A., Guerguiev, J., Bhalla, U. S., & Bhatt, D. H. (2021). Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature Neuroscience*, 24(7), 1010–1019.
- [23] Greedy, W., Zhu, H. W., Pemberton, J., Bhatt, D. H., & Richards, B. A. (2022). Single-phase deep learning in cortico-cortical networks. *Advances in Neural Information Processing Systems*, 35.
- [24] Haider, P., Ellenberger, B., Kriener, L., Jordan, J., Senn, W., & Petrovici, M. A. (2021). Latent equilibrium: A unified learning theory for arbitrarily fast computation with arbitrarily slow neurons. *Advances in Neural Information Processing Systems*, 34.

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.

- [25] Richards, B. A., & Lillicrap, T. P. (2019). Dendritic solutions to the credit assignment problem. *Current Opinion in Neurobiology*, 54, 28–36.
- [26] Hinton, G. (2022). The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345*.
- [27] Dellaferrera, G., & Bhatt, D. (2022). Error-driven input modulation: Solving the credit assignment problem without a backward pass. *International Conference on Machine Learning*, 4937–4955.
- [28] Lee, D.-H., Zhang, S., Fischer, A., & Bengio, Y. (2015). Difference target propagation. *European Conference on Machine Learning*, 498–515.
- [29] Meulemans, A., Tristany Farinha, M., García Ordoez, J., Simoncini, P., Tagliasacchi, A., & Lucchi, A. (2021). Credit assignment in neural networks through deep feedback control. *Advances in Neural Information Processing Systems*, 34.
- [30] Millidge, B., Seth, A., & Buckley, C. L. (2022). Predictive coding: A theoretical and experimental review. *arXiv preprint arXiv:2107.12979*.
- [31] Koch, C., Poggio, T., & Torre, V. (1983). Nonlinear interactions in a dendritic tree: Localization, timing, and role in information processing. *Proceedings of the National Academy of Sciences*, 80(9), 2799–2802.
- [32] Silver, R. A. (2010). Neuronal arithmetic. *Nature Reviews Neuroscience*, 11(7), 474–489.
- [33] Beniaguev, D., Segev, I., & London, M. (2021). Single cortical neurons as deep artificial neural networks. *Neuron*, 109(17), 2727–2739.

## A Supplementary Results and Figures

### Rule-Family Ranking and Broadcast-Mode Dependence

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 remains the most reliable mode across completed sweeps and robustly outperforms local-mismatch. In the completed context-gating error-shaping sweep, the best per-soma condition reaches  $0.386 \pm 0.039$  test accuracy, while local-mismatch remains at 0.114. In the focused MNIST recheck after local-mismatch stabilization, local-mismatch improves modestly for shunting but remains far below per-soma (Table 5, Fig. 14).

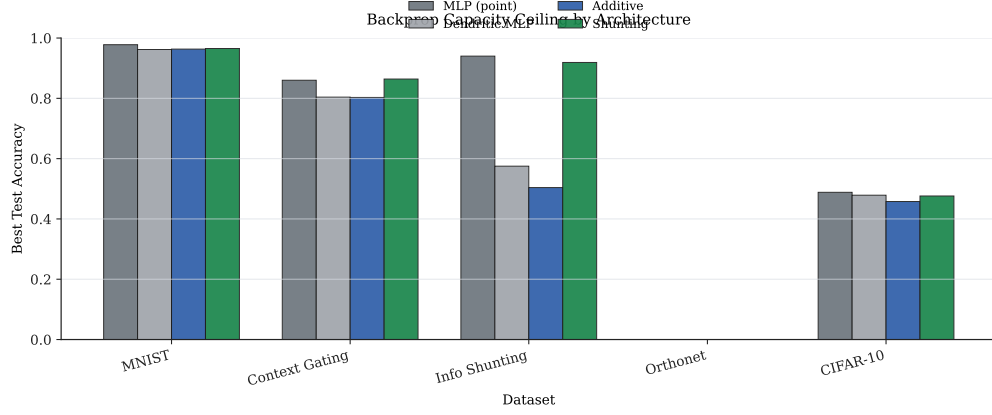


Figure 7: **Phase 1 capacity ceilings (standard backprop)**. Best standard-test accuracy per dataset and core type.

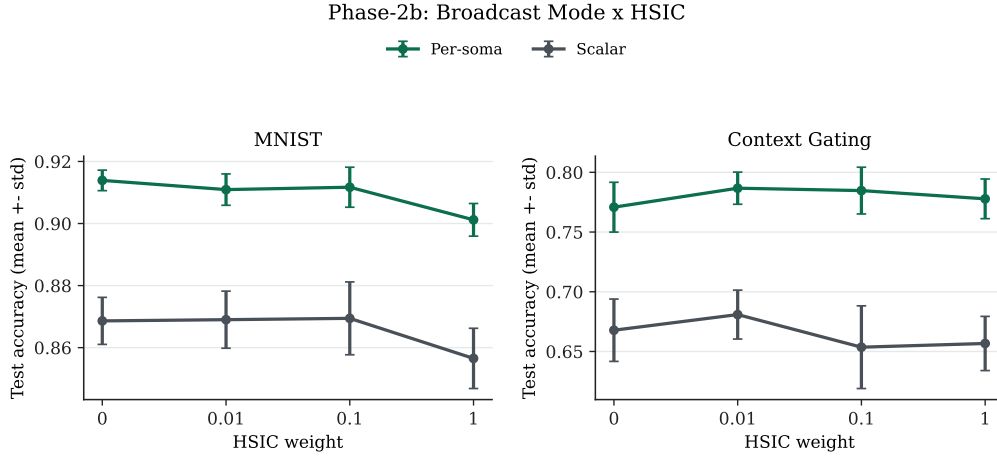


Figure 8: **HSIC strength and broadcast mode**. For context gating, moderate HSIC weights (0.01–0.1) improve local learning under per-soma broadcast, while large weights degrade performance. For MNIST, HSIC has negligible effect. Error bars:  $\pm 1$  s.d. across seeds.

#### 403 Phase-Based Sweep Figures

#### 404 Additional Gradient-Fidelity Analyses

#### 405 Small-Network Sandbox Summary

### 406 B Implementation Details (Appendix)

#### 407 Units and Normalization

#### 408 Positive Parameterization

409 To enforce  $g \geq 0$ , we use a positive parameterization (e.g.  $g = \exp(\theta)$  or softplus). For  $g = \exp(\theta)$ ,  
 410  $\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial g} \cdot g$ .

#### 411 Decoder Update Modes

412 Let  $W_{\text{dec}}$  map  $V_L \rightarrow \hat{y} \in \mathbb{R}^{d_{\text{out}}}$ . We compare: **backprop** ( $\nabla_{W_{\text{dec}}} L$  via autograd), **local** ( $\Delta W_{\text{dec}} =$   
 413  $\eta \langle \delta_0 V_L^\top \rangle_B$ ), and **frozen** ( $\Delta W_{\text{dec}} = 0$ ).

### Decoder Locality: Local Matches Backprop, None Degrades

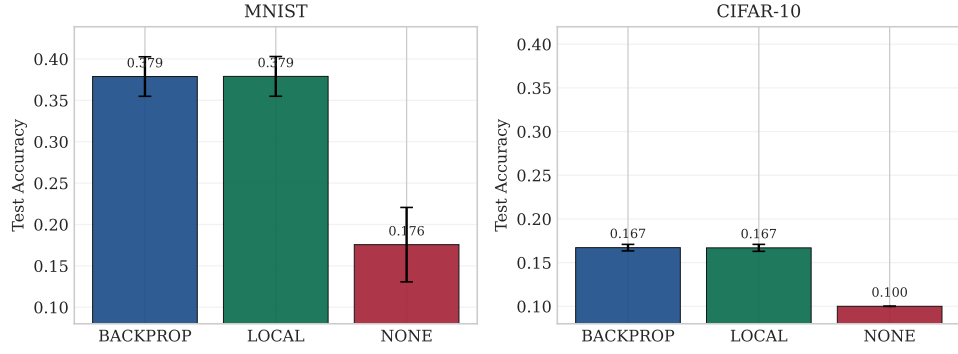


Figure 9: **Decoder locality.** On both MNIST and CIFAR-10, local decoder updates match backpropagated decoder updates, while frozen decoder weights collapse performance.

### Regime Dependence: Shunting Dominates Additive Across Inhibition Levels

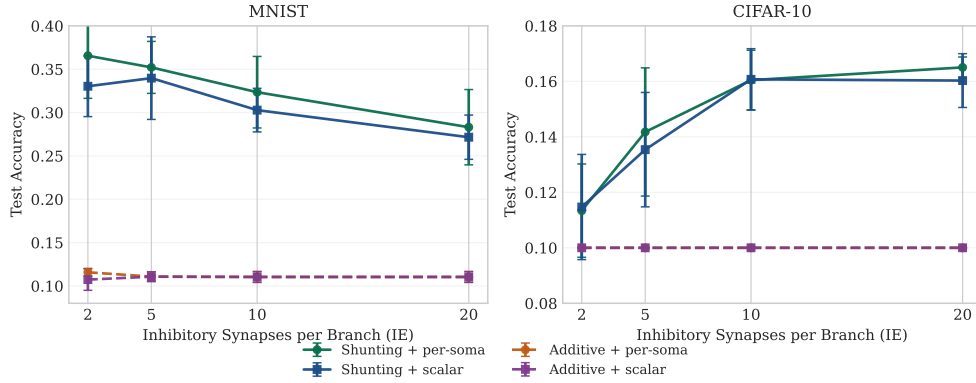


Figure 10: **Regime dependence across inhibition levels (robust sweep).** Shunting networks consistently outperform additive controls across inhibitory synapse counts and broadcast modes.

## 414 Algorithm Sketch

---

### Algorithm 1 Local Credit Assignment (schematic)

---

- 1: **Input:** Model, minibatch  $(x, y)$ , config  $\mathcal{C}$
  - 2: Forward pass; compute loss  $L$  and output error  $\delta^y = \partial L / \partial \hat{y}$
  - 3: Compute somatic error  $\delta_0 = W_{\text{dec}}^\top \delta^y$
  - 4: **for** each layer  $n$  (reverse order) **do**
  - 5:   Broadcast error:  $e_n = \text{broadcast}(\delta_0, \mathcal{C})$
  - 6:   **if** path propagation enabled **then**
  - 7:     Compute  $\pi_n$  via Eq. 26; set  $e_n \leftarrow e_n \cdot \pi_n$
  - 8:   **end if**
  - 9:   Compute layer modulators  $\rho_n$  (correlation) and  $\phi_n$  (conditional predictability)
  - 10:   Optionally convert to branch-specific factors (depth modulation, apical/basal scaling)
  - 11:   Apply local updates for synaptic and dendritic conductances (3F/4F/5F template)
  - 12: **end for**
  - 13: Clip gradients; optimizer step
-



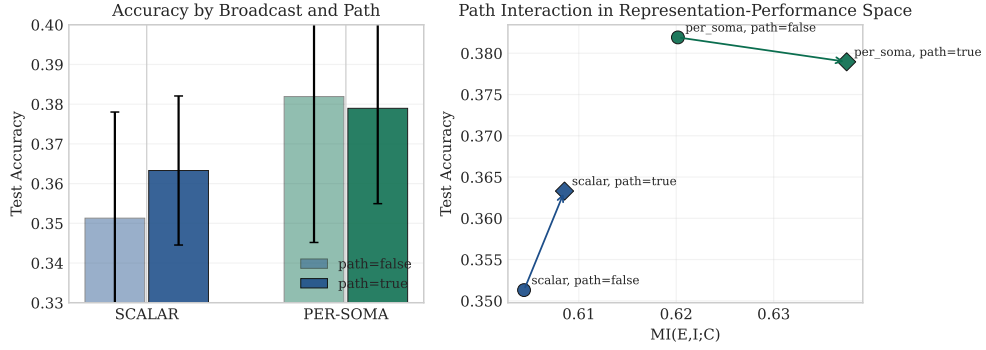


Figure 11: **Broadcast-path interaction in source analysis (robust sweep).** Within per-soma broadcast, path propagation changes information metrics more than accuracy; within scalar broadcast it improves accuracy more with smaller information gains.

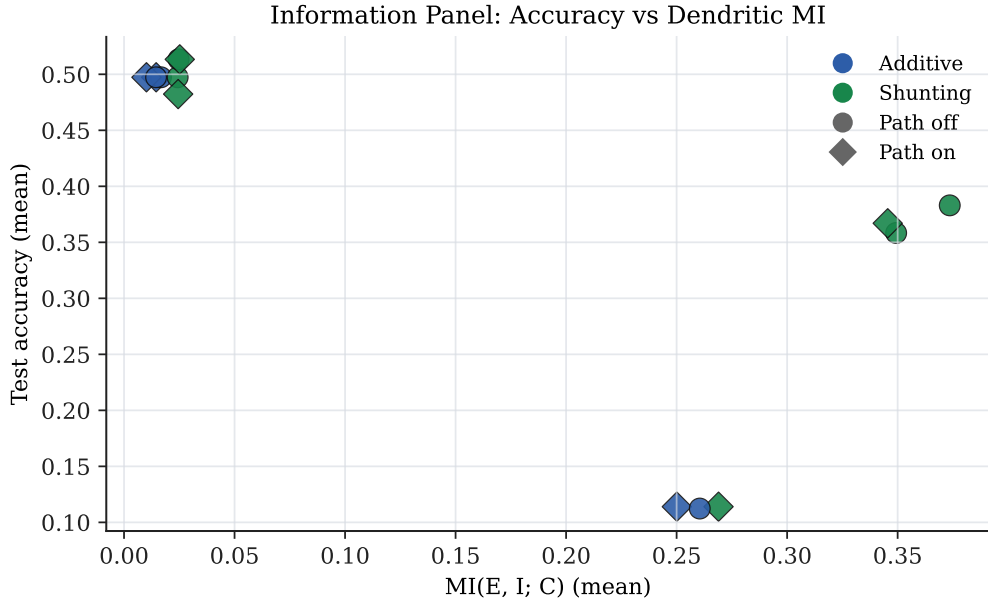


Figure 12: **Information panel.** Mutual information proxy  $I(E, I; C)$  versus test accuracy for shunting (green) and additive (blue) networks, with and without path propagation.

## 415 C Theoretical Comparison (Appendix)

## 416 D Morphology-Aware Extensions (Details)

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

419 **Path-integrated propagation.** Exact gradients in dendritic trees involve a path-sum of multiplica-  
420 tive edge factors (Eq. 10). Path propagation approximates this depth attenuation by modulating  
421 the broadcast error with the recursive path factor  $\pi_n$  (Eq. 26), implemented as a per-sample scalar  
422 broadcast within each layer.

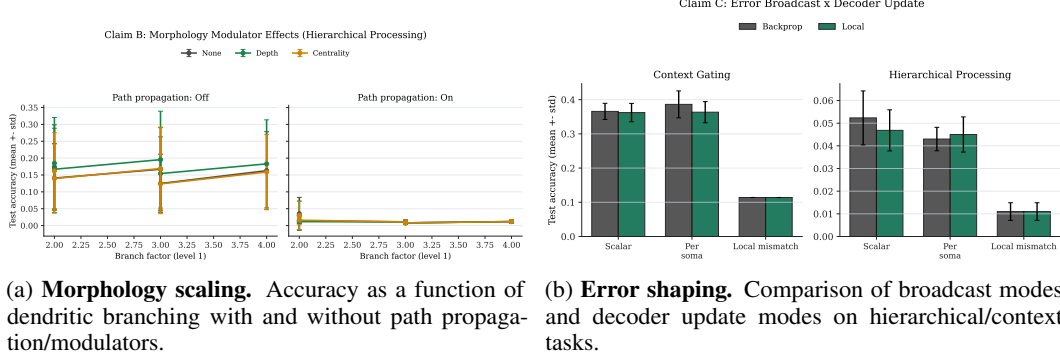


Figure 13: Phase 3 ablations: morphology and error shaping.

Core	Broadcast	Decoder	Test acc. (mean ± std)
Shunting	per-soma	local	0.9119 ± 0.0049
Shunting	per-soma	backprop	0.9091 ± 0.0075
Shunting	local-mismatch	local	0.1460 ± 0.0461
Shunting	local-mismatch	backprop	0.1457 ± 0.0370
Additive	per-soma	local	0.8938 ± 0.0073
Additive	per-soma	backprop	0.9000 ± 0.0010
Additive	local-mismatch	local	0.3418 ± 0.0575
Additive	local-mismatch	backprop	0.3475 ± 0.0952

Table 5: **Focused local-mismatch recheck (MNIST, 5F)**. Results from the completed post-fix sweep (24 configs, 3 seeds per condition). Local-mismatch no longer hard-collapses but remains substantially weaker than per-soma broadcast, especially for shunting cores.

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

**Dendritic normalization.** Normalize dendritic updates by total branch conductance:  $\Delta g_j^{\text{den}} \leftarrow \Delta g_j^{\text{den}} / (\sum_k g_k^{\text{den}} + \varepsilon)$ . This reduces update variance when total conductance  $G_n$  is large, analogous to homeostatic synaptic scaling [21].

**Apical vs basal branch differentiation.** Assign each branch a type flag  $t_j \in \{0, 1\}$  (basal, apical) 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-down learning, consistent with distinct plasticity rules in apical vs. basal dendrites of pyramidal neurons [19].

## E HSIC Auxiliary Objectives (Details)

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 HSIC losses are:

$$\mathcal{L}_{\text{HSIC}}^{\text{self}} = \frac{1}{B^2} \text{tr}(\mathbf{K}_Z \mathbf{H} \mathbf{K}_Z \mathbf{H}) \quad (\text{self-decorrelation}) \quad (29)$$

$$\mathcal{L}_{\text{HSIC}}^{\text{target}} = -\frac{1}{B^2} \text{tr}(\mathbf{K}_Z \mathbf{H} \mathbf{K}_Y \mathbf{H}) \quad (\text{target-correlation}) \quad (30)$$

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  $\partial \mathcal{L}^{\text{target}} / \partial \mathbf{Z} = -\frac{4}{B^2} \mathbf{H} \mathbf{K}_Y \mathbf{H} \mathbf{Z}$ . These are added to synaptic eligibility traces via the chain rule.

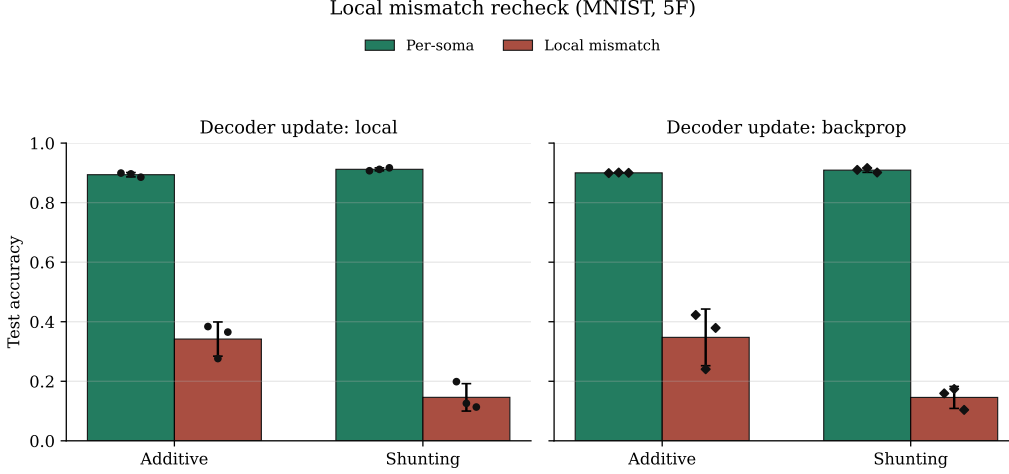


Figure 14: **Local-mismatch recheck summary.** Seed-level and mean  $\pm$  std test accuracy for per-soma vs local-mismatch under matched MNIST 5F settings. The post-fix recheck confirms modest local-mismatch recovery in shunting, but a large remaining gap to per-soma.

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

## F Online Variant with Eligibility Traces

Define continuous-time eligibilities per synapse:  $\tau_e \dot{e}_j^{\text{syn}}(t) = -e_j^{\text{syn}}(t) + x_j(t) (E_j - V_n(t)) R_n^{\text{tot}}(t)$ , and likewise for dendritic connections. With modulatory signal  $m_n(t)$ :  $\Delta g_j^{\text{syn}} \propto \int e_j^{\text{syn}}(t) m_n(t) dt$ , which instantiates three-factor learning in continuous time [17, 18].

## G Align-Then-Memorize Dynamics

Following the framework of Refinetti et al., we examine whether local learning in dendritic networks exhibits an “align-then-memorize” dynamic: an initial phase of gradient alignment followed by loss reduction. Figure 18 shows that shunting networks exhibit a characteristic U-shaped alignment trajectory—alignment initially decreases (early memorization/exploration) before recovering and stabilizing, with the alignment improvement rate transitioning from negative to positive around epoch 15–20. Additive networks show a flatter, more variable trajectory with less pronounced alignment recovery, consistent with their lower overall gradient fidelity.

## H Depth Scaling and Noise Robustness

To probe the scalability and robustness of local credit assignment, we conduct two systematic sweeps using a controlled small-network architecture (20 excitatory neurons per layer) with varying dendritic depth and error-signal noise.

**Depth scaling (Bartunov-style).** We vary the number of dendritic branch layers from 1 to 4 (branch factors [9], [3,3], [3,3,3], [3,3,3,3]) and compare standard backprop and 5F local learning on both

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 7: **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.

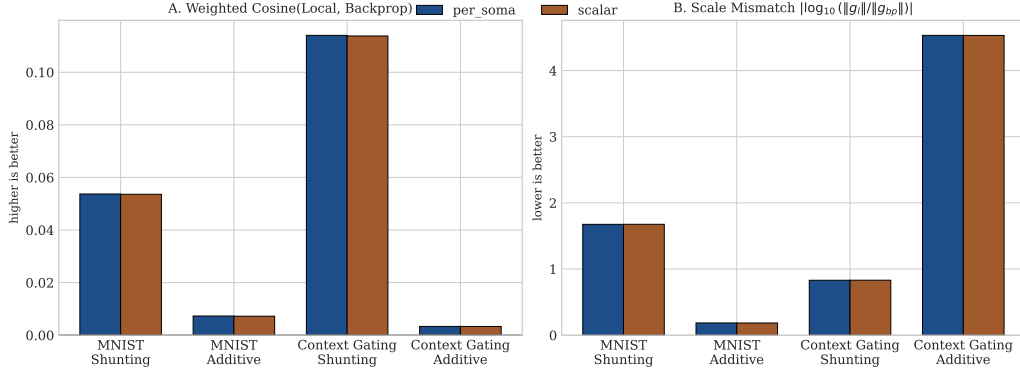


Figure 15: **Per-soma vs scalar gradient-fidelity on trained Phase-2b checkpoints.** Aggregate directional and scale metrics for shunting and additive cores on MNIST and context gating.

shunting and additive cores. Figure 19 shows accuracy versus dendritic depth (left) and the BP–local gap (right). Shunting local learning achieves 45.2% at depth 1 and degrades gracefully to 35.0% at depth 4—a gap increase from 0.43 to 0.54. Additive local learning remains near chance ( $\sim 11\%$ ) at all depths, confirming that the credit signal advantage of shunting is not an artifact of shallow architectures. Backprop accuracy is stable across depths for both core types (81–90%), with shunting consistently outperforming additive.

**Noise robustness.** We add Gaussian noise  $\mathcal{N}(0, \sigma^2)$  to the broadcast error signal during local learning, with  $\sigma \in \{0, 0.01, 0.05, 0.1, 0.5, 1.0\}$ , and compare shunting and additive cores. Figure 20 shows absolute accuracy (left) and relative accuracy normalized to the  $\sigma = 0$  baseline (right). Shunting networks are robust to small noise ( $\sigma \leq 0.05$ : 45.1% vs 45.8% baseline), degrade moderately at  $\sigma = 0.1$  (41.9%, 92% of baseline), and collapse only at extreme noise ( $\sigma \geq 0.5$ ). Additive networks remain at chance ( $\sim 10.5\%$ ) across all noise levels. This demonstrates that shunting credit signals carry genuine learning information that is progressively corrupted by noise, whereas additive networks never form useful credit signals to begin with.

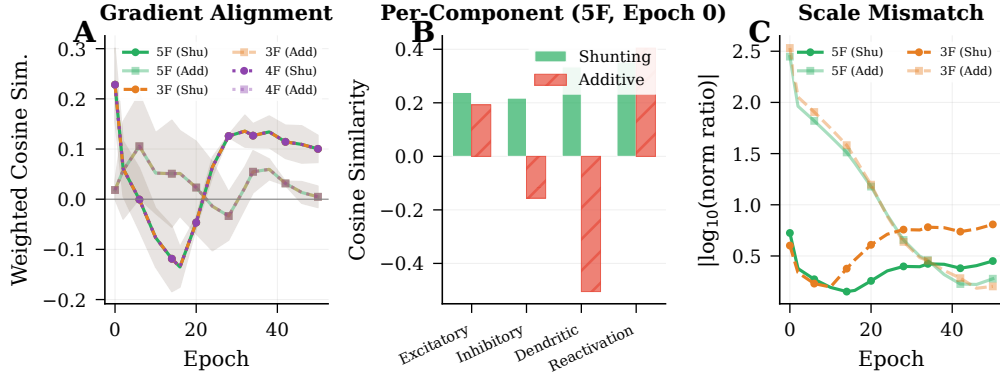


Figure 16: **Gradient fidelity over training.** (A) Parameter-count-weighted cosine similarity between local and backprop gradients over training epochs. (B) Per-component cosine at initialization. (C) Scale mismatch over training.

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 8: Units and normalization conventions.

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 conditioning than 3F; 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 9: Variant taxonomy: computational cost and current empirical regime map ( $L = \text{depth}$ ,  $d_n = \text{compartments}$ ).

Component	Biological Analog	Interpretation
Conductance scaling $R_n^{\text{tot}}$	Input resistance	Local sensitivity modulation (Lemma 1)
Driving force $(E_j - V_n)$	Synaptic current	Local gradient factor (Prop. 1)
Shunting inhibition	Divisive normalization	Sensitivity $\partial V / \partial g_{\text{inh}} \propto -V$ (Sec. 2.3)
Path factor $\pi_n$	Cable attenuation	Approximate depth attenuation (Eq. 26)
Morphology factor $\rho_n$	Layer relevance	Correlation with output (Eq. 17)
Information factor $\phi_n$	Conditional predictability	Predictability-based amplification (Eq. 23)
Dendritic normalization	Homeostasis	Stabilizes branch-scale updates (Appendix D)
Branch-type scaling	Apical vs. basal	Differential plasticity across compartments (Appendix D)
Broadcast alignment	Feedback alignment	Descent-direction in expectation (Thm. 2)

Table 10: Theoretical components and their biological/algorithmic interpretations.

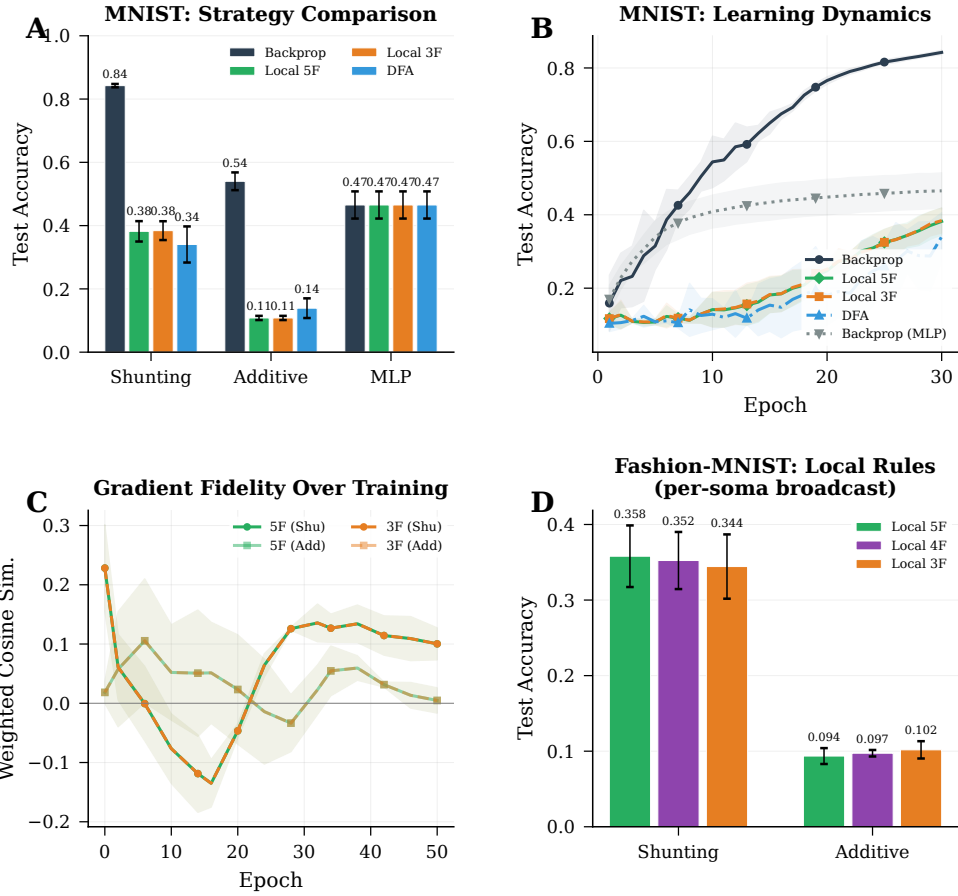


Figure 17: **Controlled small-network sandbox (summary)**. A compact view of strategy comparisons, learning dynamics, and gradient-fidelity trends in a matched small-architecture regime used for mechanism isolation.

#### Align-then-memorize analysis (Refinetti et al.)

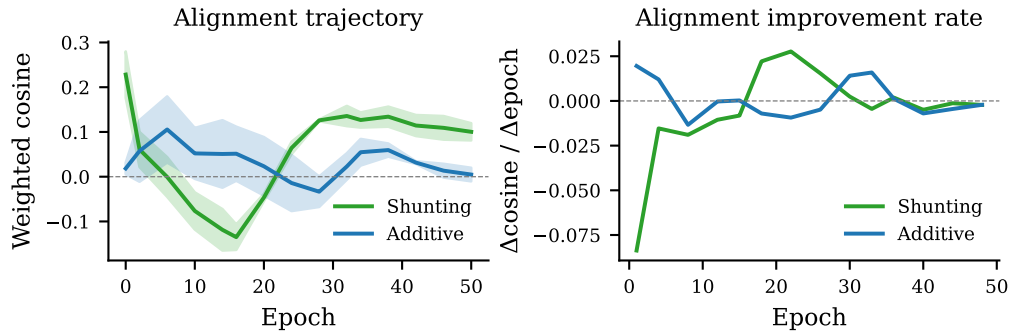


Figure 18: **Align-then-memorize analysis**. Left: alignment trajectory (weighted cosine between local and backprop gradients, averaged across layers) over training. Right: alignment improvement rate ( $\Delta \text{cosine} / \Delta \text{epoch}$ ). Shunting networks (green) show a recovery phase consistent with the align-then-memorize framework; additive networks (blue) show weaker alignment dynamics. Shaded regions:  $\pm 1$  s.e. across seeds.

### Depth scaling (Bartunov-style)

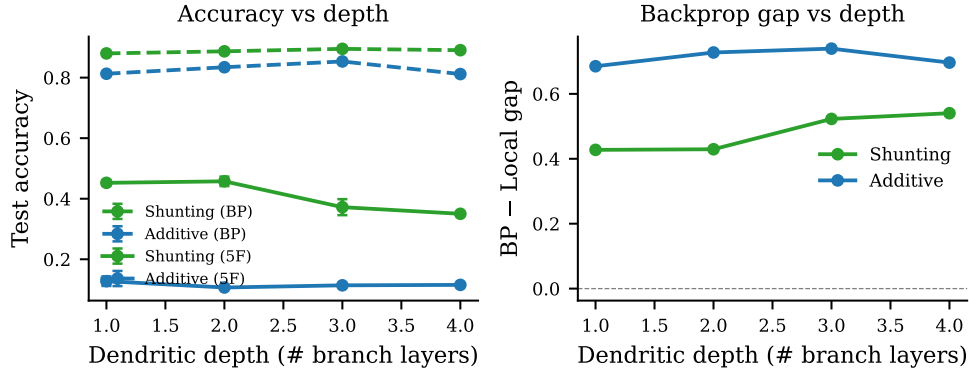


Figure 19: **Depth scaling (Bartunov-style)**. Left: test accuracy for backprop (dashed) and 5F local (solid) as a function of dendritic depth. Right: BP-local gap by depth. Shunting (green), additive (blue). Error bars:  $\pm 1$  s.e. across 3 seeds.

### Noise robustness of credit signals

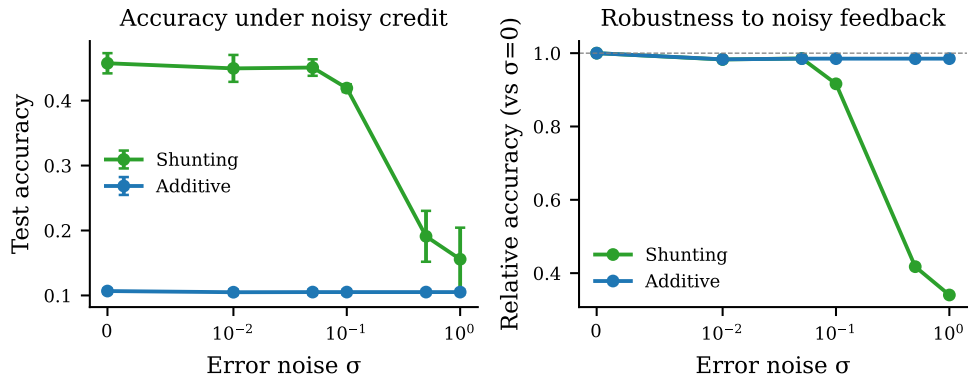


Figure 20: **Noise robustness of credit signals**. Left: test accuracy under increasing error noise  $\sigma$ . Right: relative accuracy (normalized to  $\sigma = 0$ ). Shunting (green), additive (blue). Error bars:  $\pm 1$  s.e. across 3 seeds.