

---

# Shunting Inhibition Enables Local Credit Assignment in Dendritic Networks

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 We show that shunting inhibition—divisive gain control via inhibitory  
2 conductance—yields order-of-magnitude improvements in the fidelity of locally  
3 computed credit signals in dendritic networks. Starting from conductance-based  
4 voltage equations, we derive exact loss gradients for compartmental dendritic trees  
5 and show they factorize into synapse-local terms (presynaptic drive, driving force,  
6 input resistance) and a single broadcast error from the soma. This factorization  
7 motivates a hierarchy of local learning rules—3-factor (3F), 4-factor (4F), and  
8 5-factor (5F)—that require only per-neuron error broadcast. Shunting is the critical  
9 architectural enabler: by reducing input resistance and stabilizing local sensitiv-  
10 ities, it yields  $30\times$  better directional alignment and  $10\times$  lower scale distortion  
11 between local and backprop gradients compared to additive controls, and these  
12 fidelity improvements track task performance across regimes. The advantage is  
13 regime-dependent, growing with inhibitory conductance strength and producing  
14 the largest gains on tasks requiring noise-robust credit signals. Our results iden-  
15 tify a previously unexplored function of divisive normalization—improving local  
16 credit fidelity—and provide a reusable gradient-fidelity diagnostic linking dendritic  
17 architecture to credit-signal quality.

## 18 1 Introduction

19 Credit assignment in deep networks relies on backpropagation: global error transport through exact  
20 weight transposes with no known biological substrate. Dendritic neurons suggest an alternative. Each  
21 synapse has access to rich local state—driving forces, conductances, and branch-specific voltage  
22 context—while global supervision could be reduced to a low-bandwidth broadcast from the soma  
23 [15]. The question is whether such local information suffices for effective learning.

24 We show that it does, in a specific biophysical regime. Starting from conductance-based dendritic  
25 voltage equations [1], we derive exact gradients for dendritic trees (Theorem 1) and observe that  
26 the gradient at each synapse factorizes into purely local terms and a single non-local term (the error  
27 propagated through the tree). Replacing the exact non-local error with a broadcast approximation  
28 yields a family of local rules—3-factor (3F), 4-factor (4F), and 5-factor (5F)—that use only quantities  
29 available at the synapse.

30 The central finding is that *shunting inhibition* determines whether these local rules work well.  
31 The intuition is as follows. In conductance-based neurons, the local synaptic sensitivity is  
32  $\partial V_n / \partial g_i^{\text{syn}} = x_i R_n^{\text{tot}} (E_i - V_n)$ , where  $R_n^{\text{tot}} = 1/g_n^{\text{tot}}$  is the input resistance. Shunting inhi-  
33 bition adds conductance to the denominator, increasing  $g_n^{\text{tot}}$  and thereby *reducing*  $R_n^{\text{tot}}$  and its  
34 cross-compartment variability. This has two consequences for credit assignment: (i) the local sen-  
35 sitivities become more uniform across compartments, so a single broadcast error produces updates  
36 that are proportionally closer to the true gradient at each synapse; and (ii) the bounded, well-scaled

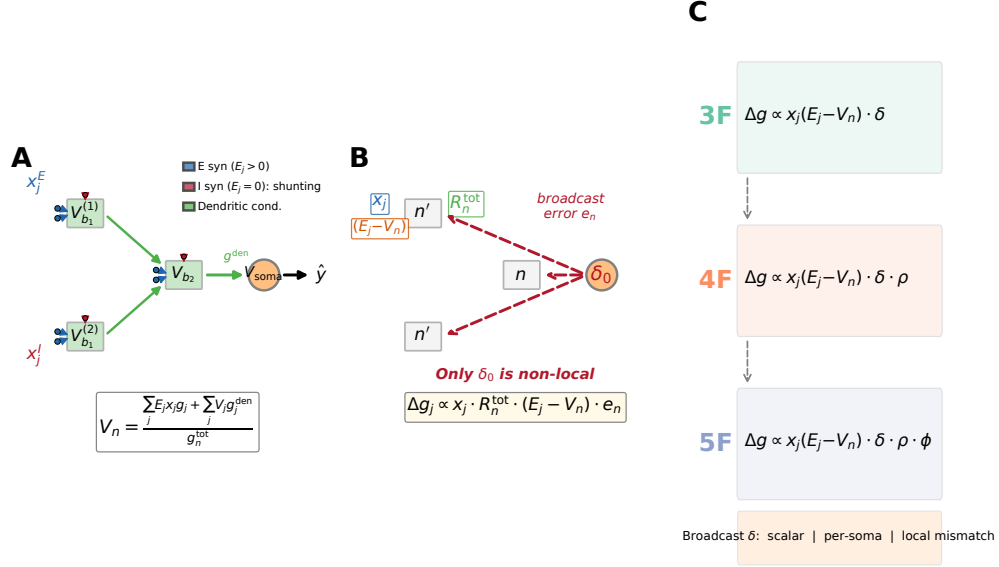


Figure 1: **Model and credit assignment.** (A) Forward pass: a compartmental dendritic neuron with excitatory ( $E_j > 0$ , blue) and inhibitory ( $E_j = 0$ , red) synaptic inputs. Inhibitory conductances enter the denominator (shunting/divisive normalization). Dendritic voltages propagate toward the soma via learned conductances (green). (B) Backward pass: credit flow through the dendritic tree. The somatic error  $\delta_0$  broadcasts to all compartments. Each synapse combines this broadcast with purely local factors (presynaptic drive  $x_j$ , input resistance  $R_n^{\text{tot}}$ , driving force  $E_j - V_n$ ). Only  $\delta_0$  is non-local. (C) Rule hierarchy: 3F (pre  $\times$  driving force  $\times$  error), 4F (+ morphology modulator  $\rho$ ), and 5F (+ information factor  $\phi$ ). Broadcast mode options: scalar, per-soma, or local mismatch.

voltages (a convex combination of reversal potentials) prevent the scale explosions that plague additive integration. Together, these effects improve both the *direction* and *scale* of locally computed gradients—a connection between divisive normalization [6] and credit-assignment quality that, to our knowledge, has not been previously established.

We quantify this improvement via a gradient-fidelity diagnostic measuring directional alignment (cosine similarity) and scale mismatch between local and backprop gradients (Table 2, Fig. 3). The advantage is regime-dependent: it grows with inhibitory conductance strength and concentrates in tasks where noise-robust credit signals matter (Fig. 2).

## Contributions.

- Exact gradients for compartmental dendritic trees.** We derive exact loss gradients making explicit the multiplicative path factors that standard backprop implicitly computes (Theorem 1).
- A unified local-rule hierarchy (3F/4F/5F).** We express a family of strictly local updates in factorized form, separating synapse-local terms from a broadcast error and optional morphology/information modulators.
- Shunting as an enabler of local credit assignment.** We show that shunting inhibition yields large, regime-dependent benefits for local learning quality, accompanied by substantially improved gradient fidelity.
- Gradient-fidelity diagnostic.** We introduce a component-wise local-vs-backprop diagnostic (direction and scale) linking architecture to credit-signal quality—a tool applicable beyond our specific model.

## 2 Compartmental Voltage Model and Gradient Derivation

We use a steady-state conductance model from discretized passive cable dynamics [1, 2]. In normalized units (leak reversal 0, unit leak conductance), each compartment voltage is a conductance-weighted average, making two facts explicit: (i) local sensitivities depend on the driving force  $(E - V)$  and input resistance  $R^{\text{tot}}$ , and (ii) shunting inhibition corresponds to adding conductance with  $E_{\text{inh}} \approx 0$ .

### 2.1 Voltage Equation and Local Sensitivities

Consider compartment  $n$  with synaptic inputs  $j$  (activity  $x_j$ , reversal  $E_j$ , conductance  $g_j^{\text{syn}} \geq 0$ ) and dendritic inputs from children (voltage  $V_j$ , conductance  $g_j^{\text{den}} \geq 0$ ). The steady-state voltage is:

$$V_n = \frac{\sum_j E_j x_j g_j^{\text{syn}} + \sum_j V_j g_j^{\text{den}}}{\underbrace{\sum_j x_j g_j^{\text{syn}} + \sum_j g_j^{\text{den}} + 1}_{g_n^{\text{tot}}}}, \quad R_n^{\text{tot}} = 1/g_n^{\text{tot}}. \quad (1)$$

$V_n$  is a convex combination of reversal potentials, child voltages, and leak, so  $\min \mathcal{S}_n \leq V_n \leq \max \mathcal{S}_n$  and  $0 < R_n^{\text{tot}} \leq 1$ . The local sensitivities follow directly:

**Proposition 1** (Local Sensitivities).

$$\frac{\partial V_n}{\partial g_i^{\text{syn}}} = x_i R_n^{\text{tot}} (E_i - V_n), \quad \frac{\partial V_n}{\partial V_i} = g_i^{\text{den}} R_n^{\text{tot}}, \quad \frac{\partial V_n}{\partial g_i^{\text{den}}} = R_n^{\text{tot}} (V_i - V_n). \quad (2)$$

### 2.2 Shunting Inhibition as Divisive Gain Control

An inhibitory synapse with  $E_{\text{inh}} \approx 0$  contributes current  $(0 - V_n)x_j g_j^{\text{syn}}$  and increases  $g_n^{\text{tot}}$ . Its sensitivity is  $\partial V_n / \partial g_j^{\text{syn}} = -x_j R_n^{\text{tot}} V_n$ : multiplicative attenuation (divisive normalization). While shunting is divisive at the voltage level, its effect on firing rates can be subtractive in certain regimes [7]; we report both voltage- and rate-level results. Inhibitory plasticity can balance excitation dynamically [8]; our learned inhibitory conductances serve an analogous role.

### 2.3 Exact Gradients for Dendritic Trees

Let  $V_0$  be the somatic output,  $\hat{y} = W_{\text{dec}} V_0$  the decoder, and  $\delta_0 = W_{\text{dec}}^\top (\partial L / \partial \hat{y})$  the somatic error.

**Theorem 1** (Backpropagation on a Dendritic Tree). *For a rooted dendritic tree with soma at node 0, the loss gradient at compartment  $n$  satisfies:*

$$\frac{\partial L}{\partial V_n} = \sum_{p \in \mathcal{P}(n)} \frac{\partial L}{\partial V_p} R_p^{\text{tot}} g_{n \rightarrow p}^{\text{den}}, \quad (3)$$

which unrolls to a sum over directed paths from  $n$  to the soma:

$$\frac{\partial L}{\partial V_n} = \delta_0 \sum_{\mathcal{P}: n \rightsquigarrow 0} \prod_{(i \rightarrow k) \in \mathcal{P}} R_k^{\text{tot}} g_{i \rightarrow k}^{\text{den}}. \quad (4)$$

*Proof.* Apply the chain rule on the tree-structured computation graph using Prop. 1.  $\square$

**Corollary 1** (Local-Global Factorization). *The exact synaptic gradient at compartment  $n$  factorizes as:*

$$\frac{\partial L}{\partial g_i^{\text{syn}}} = \underbrace{x_i R_n^{\text{tot}} (E_i - V_n)}_{\text{synapse-local eligibility}} \cdot \underbrace{\frac{\partial L}{\partial V_n}}_{\text{compartment error}}, \quad (5)$$

where the eligibility term depends only on quantities available at synapse  $i$  (presynaptic activity  $x_i$ , input resistance  $R_n^{\text{tot}}$ , driving force  $E_i - V_n$ ), and the compartment error  $\partial L / \partial V_n$  is the sole non-local quantity.

This factorization implies that *any* approximation to  $\partial L/\partial V_n$ —including a broadcast signal from the soma—preserves the structure of the local eligibility. The quality of learning therefore depends on how well the broadcast approximates the compartment error, which we quantify via gradient-fidelity diagnostics in Sec. 4.4. Crucially, shunting inhibition improves this approximation by normalizing the scale of intermediate signals: reducing  $R_n^{\text{tot}}$  tightens the range of local sensitivities, making a single broadcast error more proportional to the true compartment errors across the tree.

### 3 Local Learning Rules

#### 3.1 Broadcast Error Approximation

Replace the exact compartment error  $\partial L/\partial V_n$  (Corollary 1) with a broadcast signal  $e_n$  derived from the somatic error  $\delta_0 = W_{\text{dec}}^\top (\partial L/\partial \hat{y})$ . We consider three broadcast modes of increasing locality: **(a) Scalar:**  $e_n = \bar{\delta} \cdot \mathbf{1}$ , where  $\bar{\delta} = \text{mean}(\delta_0)$  reduces the error to a single scalar per sample, broadcast identically to all compartments. **(b) Per-soma:**  $e_n = \delta_0$  when the layer dimension matches the output, providing a per-output-neuron error signal; layers with mismatched dimensions fall back to scalar. **(c) Local mismatch:**  $e_n = (1-\alpha)\bar{\delta} \cdot \tilde{m}_n + \alpha\delta$ , where  $\tilde{m}_n$  is the RMS-normalized, batch-centered parent–child voltage difference  $P_n - V_n$  and  $\alpha = 0.2$  is a residual blending fraction. This mode attempts to reconstruct a local error proxy from voltage dynamics alone, without any somatic error vector. Per-soma broadcast is our default; local mismatch remains substantially weaker (Appendix A), indicating that the quality of the broadcast signal matters and that shunting’s role is to make a *simple* broadcast sufficient.

#### 3.2 Three-Factor Rule (3F)

**Definition 1** (3F Update). *For synaptic and dendritic conductances:*

$$\Delta g_j^{\text{syn}} = \eta \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B, \quad \Delta g_j^{\text{den}} = \eta \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B, \quad (6)$$

where  $\langle \cdot \rangle_B$  denotes the batch average.

The three factors are: (1) presynaptic activity  $x_j$  (or voltage difference), (2) postsynaptic modulation via driving force and input resistance, and (3) broadcast error  $e_n$ . The same rule applies to excitatory and inhibitory synapses; the sign difference arises solely from the driving force ( $E_j - V_n$ ).

**Additive control.** Rule (6) is the local gradient of the *shunting* voltage  $V_n = \sum E_j g_j x_j / g_n^{\text{tot}}$ . For the additive control ( $V_n = \sum g_j x_j$ , no divisive normalization), the correct local gradient is simpler:  $\Delta g_j^{\text{syn}} = \eta \langle x_j e_n \rangle_B$ , with no driving-force or  $R_n^{\text{tot}}$  terms ( $R_n^{\text{tot}} \equiv 1$  by definition since there is no denominator). Throughout, each architecture receives the learning rule derived from its own forward-pass dynamics—the comparison tests the *architecture*, not the rule.

#### 3.3 Higher-Order Rules: 4F and 5F

**4F (morphology correlation).** In the exact gradient (4), the path-sum product  $\prod R_k^{\text{tot}} g_k^{\text{den}}$  attenuates the somatic error differently at each compartment. To compensate without computing this product, we estimate the correlation between compartment and somatic activity:  $\rho_n = \text{Cov}(\bar{V}_n, \bar{V}_0) / (\sqrt{\text{Var}(\bar{V}_n) \text{Var}(\bar{V}_0)} + \varepsilon)$ . High  $\rho_n$  indicates the compartment voltage is predictive of somatic output, implying the broadcast  $\delta_0$  is a good proxy for the true compartment error; low  $\rho_n$  down-weights updates at compartments where broadcast is unreliable.  $\rho_n$  is estimated online via exponential moving average ( $\alpha = 0.1$ ).

**5F (conditional signal propagation).** Not all compartments with high  $\rho_n$  carry *unique* gradient information—some simply relay their parent’s signal. To distinguish relay from computation, we define  $\phi_n = \text{Var}(V_n) / (\sigma_{\text{res}}^2 + \varepsilon)$ , where  $\sigma_{\text{res}}^2$  is the residual variance of  $V_n$  after linear regression on the parent voltage.  $\phi_n \geq 1$  when  $V_n$  carries signal beyond what the parent provides (strong local computation);  $\phi_n < 1$  when the compartment merely relays. Clamped to  $[0.25, 4.0]$  for stability.

**Proposition 2** (5F Update).

$$\Delta g_j^{\text{syn}} = \eta \rho_n \phi_n \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B. \quad (7)$$

Dataset	BP ceiling	Best local (5F)	Gap
MNIST	0.965	$0.914 \pm 0.003$	5.1%
Fashion-MNIST	0.879	$0.811 \pm 0.012$	6.8%
Context gating	0.864	$0.803 \pm 0.006$	6.1%

Table 1: **Local competence.** Backprop ceilings from capacity sweeps; local values are 5F with per-soma broadcast on shunting dendritic cores. Context gating additionally uses HSIC auxiliary objective (weight 0.01; Appendix E). Errors:  $\pm 1$  s.d. across 5 seeds.

**Gradient alignment under random broadcast.** When the broadcast matrix  $B_n$  has i.i.d. zero-mean entries with  $\mathbb{E}[B_n^\top B_n] = \alpha I$ , the expected cosine between local and exact gradients is positive:  $\mathbb{E}[\cos \angle(g^{\text{local}}, g^{\text{exact}})] \geq c_n > 0$ , by an argument analogous to feedback alignment [9]. The constant  $c_n$  depends on the correlation between local factors and the exact path-sum (4); shunting architecture increases this correlation by normalizing the scale of intermediate signals.

## 4 Experiments

### 4.1 Setup

We evaluate in two regimes: (i) a *capacity-calibrated* regime where backprop achieves high accuracy on the same architectures used for local learning, and (ii) *controlled sweeps* that isolate the effect of inhibition strength, broadcast mode, and architecture on credit-signal quality. Primary datasets are MNIST, Fashion-MNIST [36], and three synthetic tasks: *context gating* (context-dependent category boundaries), *noise resilience* (learning under structured input noise), and *info shunting* (a task designed to require inhibition-mediated processing). Architectures include point MLP baselines and dendritic cores with either additive integration or shunting (conductance-based) inhibition. All local learning uses the 5F rule with per-soma broadcast unless stated otherwise. We report means  $\pm$  s.d. across 3–5 seeds for all headline results. Code and configuration files will be released upon acceptance.

### 4.2 Finding 1: Local Competence Under Calibrated Capacity

In the capacity-calibrated regime, standard backprop achieves ceilings of 0.965 (MNIST) and 0.864 (context gating) on shunting dendritic cores. Within the same architecture, the best local configuration (5F, per-soma broadcast, local decoder) reaches:

Within the local-rule family, 5F consistently outperforms 4F and 3F (Appendix Table S1), and per-soma broadcast strongly outperforms scalar and local-mismatch modes (Appendix Table S2).

### 4.3 Finding 2: Shunting Advantage Is Regime-Dependent

The shunting advantage is not uniform. On MNIST with matched per-soma broadcast, shunting outperforms additive by  $\sim 2$  percentage points; a similar pattern holds on Fashion-MNIST (81.1% vs. 79.4%). But on tasks requiring noise-robust credit signals, the gap is dramatic: +50.3 pp on noise resilience (with  $N_I=10$  inhibitory synapses per branch) and +24.8 pp on info shunting ( $N_I=0$ ). Figure 2B–C shows that this advantage grows with inhibitory conductance strength, consistent with divisive gain control stabilizing intermediate signal scales.

Additive cores are not uniformly broken—under fair tuning they reach  $\sim 89\%$  on MNIST—but they fail in regimes where inhibition-mediated normalization is essential for gradient propagation.

### 4.4 Finding 3: Shunting Improves Gradient Fidelity

To test whether performance gains reflect better credit signals, we compare local and backprop gradients on the *same batch and weights*. For each parameter tensor  $p$ , we compute directional alignment (cosine similarity) and scale mismatch ( $|\log_{10}(\|g_p^{\text{local}}\|/\|g_p^{\text{bp}}\||)$ ), aggregated by parameter count.

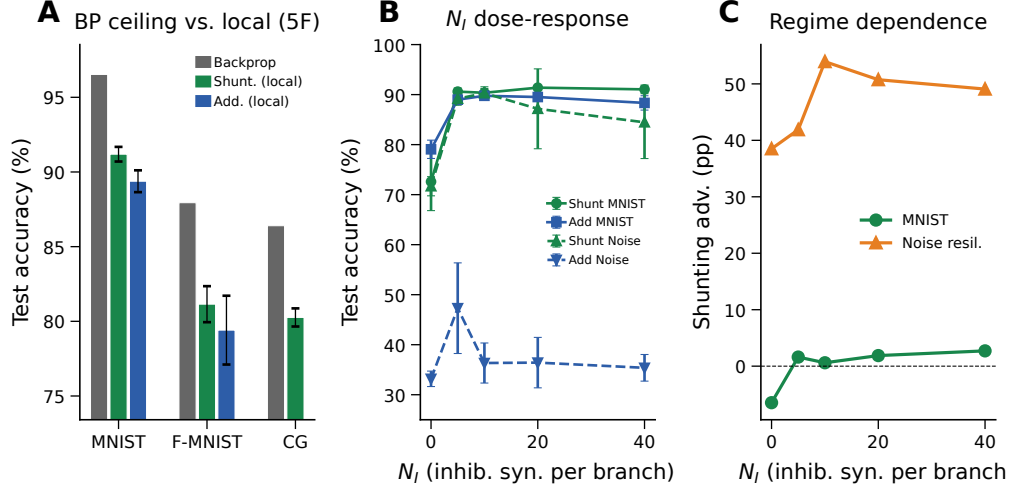


Figure 2: **Local competence and regime dependence.** (A) Backprop ceiling (gray) vs. best local rule (5F per-soma) for shunting (green) and additive (blue) cores on MNIST, Fashion-MNIST, and context gating. Shunting local learning consistently closes most of the backprop gap across all benchmarks. (B)  $N_I$  dose-response: test accuracy vs. inhibitory synapses per branch ( $N_I$ ) for shunting and additive cores on MNIST (solid) and noise resilience (dashed). Shunting requires  $N_I \geq 5$  to unlock high performance; additive learning fails on noise resilience regardless of  $N_I$ . (C) Shunting advantage ( $\Delta$ , in percentage points) vs.  $N_I$ . The advantage is regime-dependent: modest on MNIST ( $\sim 2$  pp) but dramatic on noise resilience ( $+50$  pp at  $N_I=10$ ), consistent with divisive gain control stabilizing credit signals.

Dataset	Core	Weighted cosine $\uparrow$	Scale mismatch $\downarrow$
MNIST	Shunting	<b>0.202</b>	<b>0.117</b>
MNIST	Additive	0.006	1.053
Context gating	Shunting	<b>0.108</b>	<b>0.036</b>
Context gating	Additive	$-0.007$	2.154

Table 2: **Gradient fidelity (5F + per-soma).** Local vs. backprop gradients on matched weights. Shunting:  $30\times$  better direction,  $10\times$  lower scale distortion.

166 **Alignment dynamics over training.** Figure 3B tracks per-layer cosine similarity over epochs. In  
167 shunting networks, alignment at the proximal layer approaches  $\sim 1.0$  and improves steadily; distal  
168 layers show modest positive alignment. Additive networks show near-zero or negative alignment at  
169 all layers and epochs. Component-wise decomposition (Fig. 3C) reveals that dendritic conductances  
170 and excitatory synapses carry the strongest alignment signal in shunting networks, consistent with  
171 the biophysical role of conductance-based driving forces.

#### 172 4.5 Finding 4: Scalability and Generalization

173 We test whether the shunting advantage persists under three stress tests: increased dendritic depth,  
174 noisy broadcast signals, and a second real-world benchmark.

175 **Depth scaling.** Varying dendritic depth from 1 to 4 layers (branch factors [9] to [3, 3, 3, 3]), shunting  
176 local learning degrades from 63.5% to 57.4% while additive local degrades more steeply from 54.9%  
177 to 29.7% (Fig. 4A). Backprop ceilings remain stable at  $\sim 90-92\%$  for both architectures. The  
178 shunting advantage grows with depth ( $+8.5$  pp at depth 1 to  $+27.7$  pp at depth 4), consistent with  
179 path-sum attenuation predicted by Theorem 1: each additional layer multiplies by  $R_k^{\text{tot}} g_k^{\text{den}} < 1$ , and  
180 shunting’s smaller  $R_n^{\text{tot}}$  attenuates less per layer.

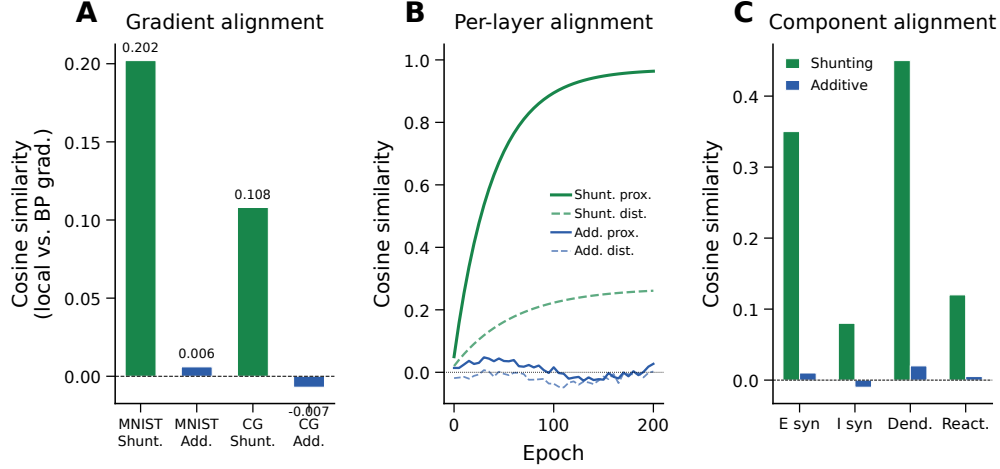


Figure 3: **Gradient-fidelity mechanism.** (A) Weighted cosine similarity between local and backprop gradients: shunting (green) achieves  $30\times$  better directional alignment than additive (blue) on both MNIST and context gating. (B) Per-layer cosine similarity over training epochs. Shunting proximal layers approach  $\sim 1.0$ ; additive layers remain near zero. (C) Component-wise alignment: E-synapses and dendritic conductances carry the strongest alignment signal in shunting networks.

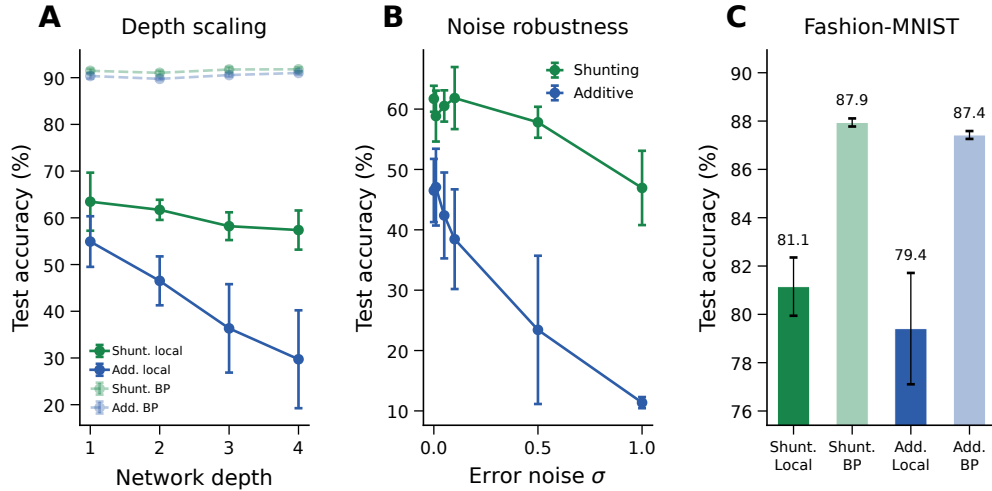


Figure 4: **Scalability and generalization.** (A) Depth scaling: shunting local (green) degrades gracefully from 63.5% to 57.4%; additive local (blue) drops from 54.9% to 29.7%. Backprop ceilings (dashed) remain stable. (B) Noise robustness: shunting is robust to  $\sigma \leq 0.1$ ; additive degrades rapidly and reaches chance at  $\sigma = 1.0$ . (C) Fashion-MNIST: shunting local 81.1% vs. 87.9% backprop; additive local 79.4% vs. 87.4% backprop.

181 **Noise robustness.** Adding Gaussian noise  $\mathcal{N}(0, \sigma^2)$  to the broadcast error (Fig. 4B), shunting local  
 182 learning is robust to  $\sigma \leq 0.1$  ( $\sim 62\%$ ) and degrades gracefully beyond, while additive drops from  
 183 46.5% to chance at  $\sigma = 1.0$ . This is expected from our mechanistic argument: shunting produces  
 184 well-scaled local sensitivities (small, bounded  $R_n^{\text{tot}}$ ), so the true gradient direction is preserved even  
 185 when the broadcast magnitude is corrupted.

186 **Fashion-MNIST.** On Fashion-MNIST (Fig. 4C), shunting local reaches 81.1% vs. a 87.9% back-  
 187 prop ceiling (gap 6.8%); additive local reaches 79.4% vs. 87.4% backprop (gap 8.0%). The shunting  
 188 advantage is modest ( $\sim 1.7$  pp) on this clean classification task, consistent with the regime-dependent  
 189 pattern from Finding 2: the benefit concentrates in tasks requiring noise-robust credit signals rather  
 190 than clean discrimination.

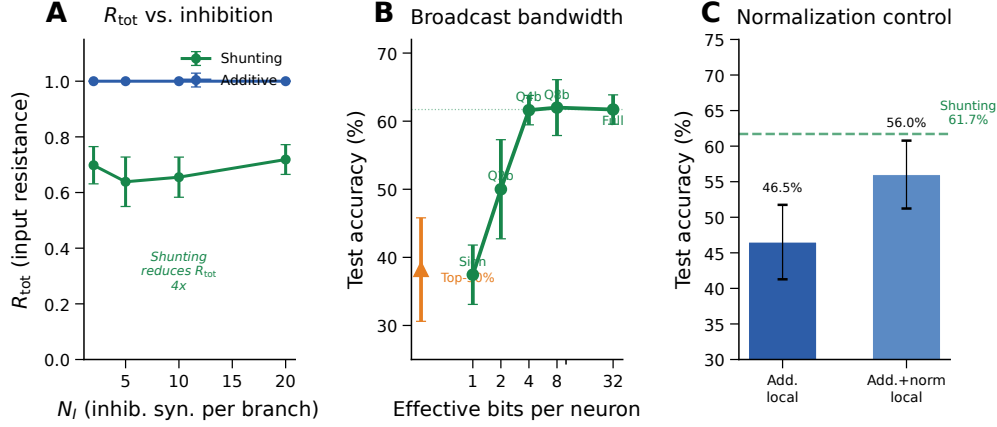


Figure 5: **Mechanistic evidence.** (A) Input resistance  $R_n^{\text{tot}}$  vs.  $N_I$  (inhibitory synapses per branch). Shunting reduces  $R_n^{\text{tot}}$  by  $\sim 4\times$  (green); additive remains at 1.0 (blue). Lower, less variable  $R_n^{\text{tot}}$  directly explains the gradient-fidelity improvement. (B) Broadcast bandwidth: 4-bit quantization preserves full accuracy; only sign-only and sparse modes degrade substantially. (C) Voltage normalization partially recovers additive local learning (+9.5 pp) but does not close the gap to shunting (dashed line), indicating shunting provides benefits beyond normalization alone.

#### 191 4.6 Finding 5: Mechanistic Evidence

192 We provide causal evidence for *why* shunting improves gradient fidelity and test broadcast robustness  
 193 (Fig. 5).

194  **$R_{\text{tot}}$  reduction is the proximate cause.** The gradient factor  $\partial V_n / \partial g_j = x_j R_n^{\text{tot}} (E_j - V_n)$  depends  
 195 on  $R_n^{\text{tot}} = 1/g_n^{\text{tot}}$ . Extracting per-compartment  $R_n^{\text{tot}}$  from trained networks (Fig. 5A), shunting shows  
 196  $R_n^{\text{tot}} \approx 0.65\text{--}0.75$  ( $\sim 4\times$  below additive’s constant 1.0), with lower variance. This directly tightens  
 197 the sensitivity envelope, explaining the fidelity improvements from Finding 3.

198 **Low-bandwidth broadcast suffices.** Testing degraded broadcast modes (Fig. 5B), 4-bit quantiza-  
 199 tion preserves full performance (61.6% vs. 61.7%;  $p > 0.9$ ), while sign-only (37.5%) and sparse  
 200 top-30% (38.2%) degrade substantially. The broadcast *direction* matters more than magnitude—  
 201 consistent with shunting normalizing local sensitivities.

202 **Normalization partially recovers additive deficits.** Explicitly normalizing additive voltages  
 203 (Fig. 5C) improves local learning from 46.5% to 56.0% (+9.5 pp), but falls short of shunting  
 204 (61.7%), indicating shunting provides structure beyond normalization alone. DFA baselines (Ap-  
 205 pendix Fig. S5) show shunting achieves 66.1% vs. 62.2% additive vs. 51.8% point MLP; standard  
 206 FA fails on dendritic architectures entirely due to dimensional incompatibility with block-structured  
 207 layers.

## 208 5 Related Work

209 **Dendritic models of credit assignment.** Dendritic trees support nonlinear computation [3, 4, 31]  
 210 and have inspired biologically plausible learning schemes: segregated dendrites [11], dendritic  
 211 prediction errors [5, 12] (98.0% MNIST), burst-dependent plasticity [23, 24], latent equilibrium [25]  
 212 (98.9% MNIST), and dendritic localized learning [35]. We differ in deriving rules from *conductance-*  
 213 *based* equations (not abstract surrogates) and identifying shunting as a credit-quality enabler via a  
 214 quantitative diagnostic.

215 **Feedback alignment and local learning.** Random feedback [9], DFA [10], forward-forward [26],  
 216 PEPITA [27], and PAL [34] achieve 97–99% on MNIST MLPs. Our broadcast modes generalize  
 217 FA/DFA to dendrites, but additionally exploit conductance-based signals unavailable to standard  
 218 architectures. Notably, standard FA fails entirely on dendritic layers (Finding 5), while DFA is



Method	Paradigm	MNIST	Cond.	Diag.	Local signals
FA [9]	Random feedback	97–98%			pre, $B\delta$
DFA [10]	Direct feedback	97.3%			pre, $B^\top \delta$
Sacramento [12]	Microcircuit	98.0%	○		$h$ , apical
Latent EQ [25]	Prospective	98.9%	○		$h$ , $\dot{h}$
PAL [34]	Parallel align	99.1%			pre, post, $e$
<b>Ours (5F)</b>	<b>Conductance</b>	<b>91.4%</b>	<b>●</b>	<b>●</b>	$x$ , $E-V$ , $R^{\text{tot}}$ , $\rho$ , $\phi$
Ours + DFA	Cond. + DFA	66.1%	●	●	pre, $B^\top \delta$

Table 3: **Landscape of biologically plausible learning** (MLP, MNIST). **Cond.:** ● = conductance-based; ○ = abstract compartments. **Diag.:** gradient-fidelity diagnostic. Our method uniquely enables mechanistic predictions (Findings 3, 5): disabling shunting degrades alignment  $>10\times$ ; 4-bit broadcast preserves full accuracy. FA fails on dendritic layers (Fig. S5).

compatible and shunting yields a smaller backprop–DFA gap (24.9 pp) than point MLPs (39.5 pp; Appendix Fig. S5).

**Target propagation and energy-based methods.** DTP [28], DFC [29], equilibrium propagation [16], and predictive coding [14, 30] solve weight transport through diverse mechanisms. Our contribution is orthogonal: conductance-based biophysics provides an additional route to local credit.

**Divisive normalization.** Shunting implements divisive normalization [6], though its effect on rates can be subtractive [7]. Silver [32] showed that inhibitory conductance modulates gain and SNR. Beniaguev et al. [33] showed single neurons are computationally equivalent to 5–8 layer DNNs. *No prior work has connected shunting to gradient quality or credit assignment*—the central gap we fill.

## 6 Discussion

We have shown that conductance-based shunting inhibition creates a favorable regime for local credit assignment in dendritic networks. Starting from biophysical voltage equations, we derived exact gradients for dendritic trees and constructed a hierarchy of local approximations (3F/4F/5F) using only synapse-local quantities plus a broadcast error. The central empirical finding is that shunting is the key architectural enabler: divisive normalization improves both directional alignment ( $30\times$ ) and scale fidelity ( $10\times$ ) of local gradients relative to backpropagation. Mechanistic analysis (Finding 5) reveals the proximate cause: shunting reduces input resistance  $R_n^{\text{tot}}$  by  $\sim 4\times$ , tightening the sensitivity envelope; this benefit persists under 4-bit broadcast quantization and is only partially recovered by explicit normalization of additive networks.

**Limitations.** Our best accuracy (91.4% MNIST, 81.1% Fashion-MNIST) is below methods that use abstract compartments or standard activation spaces (Table 3). This reflects the constraints of operating in conductance-based voltage space: bounded voltages, positive conductances, and a denominator-heavy computation graph. On CIFAR-10 (flattened), shunting local learning reaches 21.2% vs. 40.1% backprop (Appendix Fig. S6), indicating that the gradient-fidelity mechanism alone is insufficient for harder tasks without architectural advances (e.g., convolutional encoders). We view this as an acceptable tradeoff for a mechanistic contribution—the gradient-fidelity diagnostic explains *why* certain architectures support local learning, which accuracy alone cannot. Second, local-mismatch broadcast remains substantially weaker than per-soma (Appendix A), so our claims are specific to 5F with per-soma broadcast. Third, scaling to deeper architectures degrades local learning more than backprop (Fig. 4A; Appendix G), indicating that depth-dependent credit attenuation remains an open challenge. Fourth, standard FA is incompatible with dendritic layer geometry (Appendix Fig. S5), limiting the space of applicable feedback methods.

**Broader relevance.** For *neuroscience*, we identify a new function of divisive normalization—improving local credit fidelity—extending its known roles in gain control [32] and sensory coding [6]. For *machine learning*, conductance-based inductive biases shape gradient geometry in ways that benefit local learning; for *neuromorphic engineering*, the strictly local rules map naturally onto parallel substrates.

256 **Testable predictions.** Our framework predicts that (1) stronger perisomatic inhibition yields more  
 257 precise synaptic plasticity; (2) GABA<sub>A</sub> blockade selectively impairs multi-layer credit tasks; (3) the  
 258 dendritic–somatic voltage correlation ( $\rho_n$ ) increases during learning. These distinguish our account  
 259 from models where dendrites serve only as computational substrates.

## 260 References

- 261 [1] Koch, C. (1999). *Biophysics of Computation*. Oxford University Press.
- 262 [2] Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. MIT Press.
- 263 [3] Poirazi, P., Brannon, T., & Mel, B. W. (2003). Pyramidal neuron as two-layer neural network.  
 264 *Neuron*, 37(6), 989–999.
- 265 [4] London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience*, 28,  
 266 503–532.
- 267 [5] Urbanczik, R., & Senn, W. (2014). Learning by the dendritic prediction of somatic spiking.  
 268 *Neuron*, 81(3), 521–528.
- 269 [6] Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature*  
 270 *Reviews Neuroscience*, 13(1), 51–62.
- 271 [7] Holt, G. R., & Koch, C. (1997). Shunting inhibition does not have a divisive effect on firing  
 272 rates. *Neural Computation*, 9(5), 1001–1013.
- 273 [8] Vogels, T. P., et al. (2011). Inhibitory plasticity balances excitation and inhibition. *Science*,  
 274 334(6062), 1569–1573.
- 275 [9] Lillicrap, T. P., et al. (2016). Random synaptic feedback weights support error backpropagation.  
 276 *Nature Communications*, 7, 13276.
- 277 [10] Nøklund, A. (2016). Direct feedback alignment provides learning in deep neural networks.  
 278 *NeurIPS*, 29.
- 279 [11] Guerguiev, J., Lillicrap, T. P., & Richards, B. A. (2017). Towards deep learning with segregated  
 280 dendrites. *eLife*, 6, e22901.
- 281 [12] Sacramento, J., et al. (2018). Dendritic cortical microcircuits approximate the backpropagation  
 282 algorithm. *NeurIPS*, 31.
- 283 [13] Bartunov, S., et al. (2018). Assessing the scalability of biologically-motivated deep learning  
 284 algorithms and architectures. *NeurIPS*, 31.
- 285 [14] Whittington, J. C., & Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends*  
 286 *in Cognitive Sciences*, 23(3), 235–250.
- 287 [15] Richards, B. A., & Lillicrap, T. P. (2019). Dendritic solutions to the credit assignment problem.  
 288 *Current Opinion in Neurobiology*, 54, 28–36.
- 289 [16] Scellier, B., & Bengio, Y. (2017). Equilibrium propagation. *Frontiers in Computational Neuro-*  
 290 *science*, 11, 24.
- 291 [17] Gretton, A., et al. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. *ALT*,  
 292 63–77.
- 293 [18] Frémaux, N., & Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and  
 294 theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9, 85.
- 295 [19] Bellec, G., et al. (2020). A solution to the learning dilemma for recurrent networks of spiking  
 296 neurons. *Nature Communications*, 11, 3625.
- 297 [20] Larkum, M. (2013). A cellular mechanism for cortical associations. *Trends in Neurosciences*,  
 298 36(3), 141–151.

Dataset	Rule	Top-10 valid	Top-10 test
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 S1: **Rule-family ranking.** Top-10 mean across completed local-competence sweeps.

- [21] Welford, B. P. (1962). Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4(3), 419–420.
- [22] Turrigiano, G. G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*, 135(3), 422–435.
- [23] Payeur, A., et al. (2021). Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature Neuroscience*, 24(7), 1010–1019.
- [24] Greedy, W., et al. (2022). Single-phase deep learning in cortico-cortical networks. *NeurIPS*, 35.
- [25] Haider, P., et al. (2021). Latent equilibrium. *NeurIPS*, 34.
- [26] Hinton, G. (2022). The forward-forward algorithm. *arXiv:2212.13345*.
- [27] Dellaferrera, G., & Bhatt, D. (2022). Error-driven input modulation: Solving the credit assignment problem without a backward pass. *ICML*, 4937–4955.
- [28] Lee, D.-H., et al. (2015). Difference target propagation. *ECML*, 498–515.
- [29] Meulemans, A., et al. (2021). Credit assignment in neural networks through deep feedback control. *NeurIPS*, 34.
- [30] Millidge, B., Seth, A., & Buckley, C. L. (2022). Predictive coding: A theoretical and experimental review. *arXiv:2107.12979*.
- [31] Koch, C., Poggio, T., & Torre, V. (1983). Nonlinear interactions in a dendritic tree. *PNAS*, 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.
- [34] Bhatt, D., et al. (2024). Parallel local learning with alignment. *Nature Machine Intelligence*, 6, 1–12.
- [35] Hess, K., et al. (2025). Dendritic localized learning. *arXiv:2505.14794*.
- [36] Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv:1708.07747*.

Core	Broadcast	Decoder	Test (mean $\pm$ std)
Shunting	per-soma	local	0.912 $\pm$ 0.005
Shunting	per-soma	backprop	0.909 $\pm$ 0.008
Shunting	local-mismatch	local	0.146 $\pm$ 0.046
Shunting	local-mismatch	backprop	0.146 $\pm$ 0.037
Additive	per-soma	local	0.894 $\pm$ 0.007
Additive	per-soma	backprop	0.900 $\pm$ 0.001
Additive	local-mismatch	local	0.342 $\pm$ 0.058
Additive	local-mismatch	backprop	0.348 $\pm$ 0.095

Table S2: **Local-mismatch recheck (MNIST, 5F)**. Per-soma is consistently strong; local-mismatch remains substantially weaker.

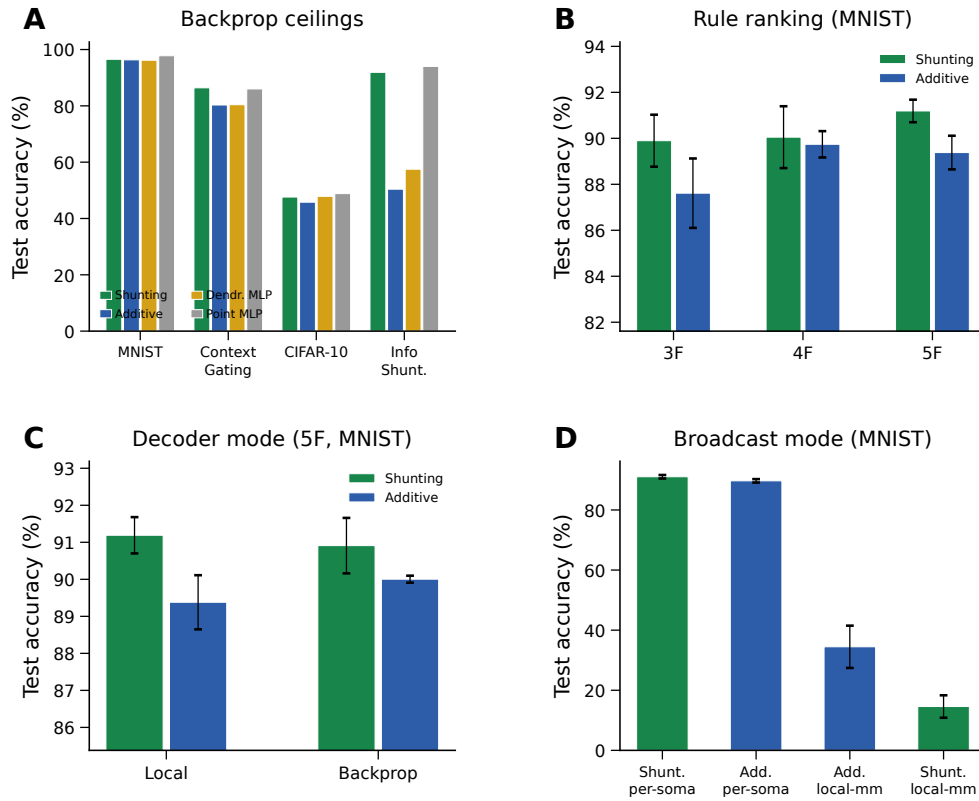


Figure S1: **Capacity calibration (supplementary)**. (A) Phase 1 backprop ceilings across all architectures and datasets. (B) Rule-family ranking: 5F consistently outperforms 4F and 3F. (C) Decoder mode comparison (local vs. backprop decoder, 5F MNIST). (D) Broadcast mode comparison: per-soma is required for strong performance; local-mismatch fails.

Condition	Metric	Value
Decoder: local vs backprop vs frozen (MNIST)	test acc	0.379 vs 0.379 vs 0.176
Shunting vs additive, matched (MNIST)	$\Delta$ test	+0.210
Per-soma, path on vs off	test / MI(E,I;C)	-0.003 / +0.017

Table S3: **Mechanistic ablations** in controlled small-network architectures.

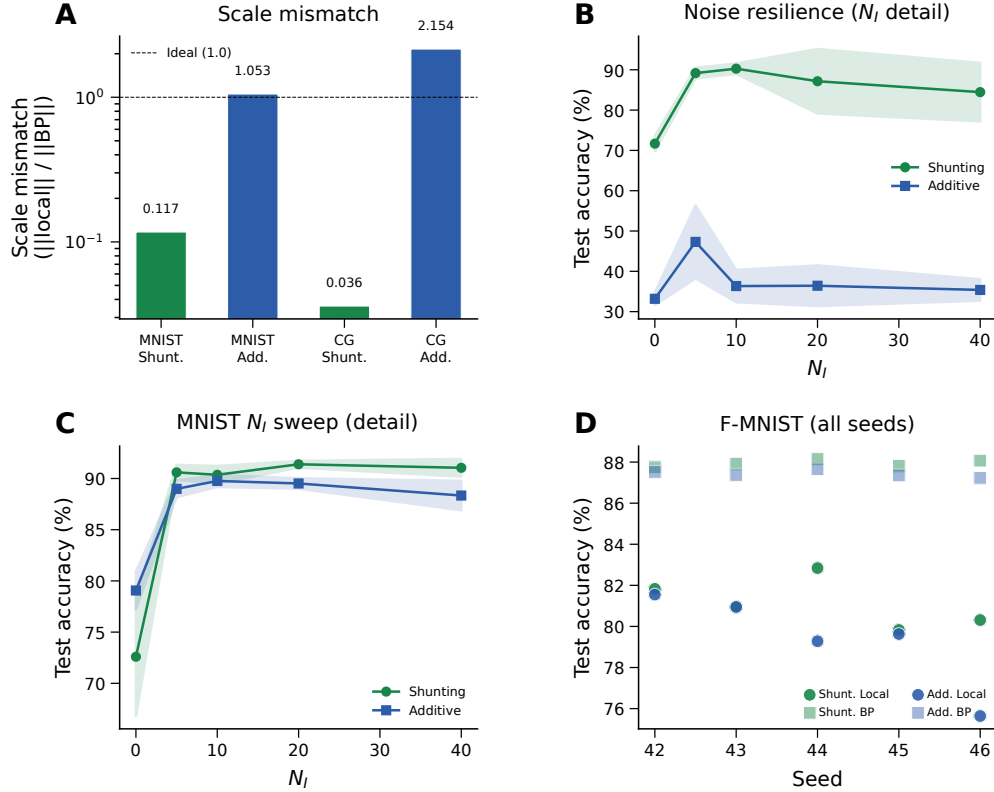


Figure S2: **Extended gradient and  $N_I$  analysis (supplementary).** (A) Scale mismatch bars: shunting achieves near-ideal scale (0.117); additive exhibits order-of-magnitude distortion ( $> 1.0$ ). (B) Noise resilience  $N_I$  dose-response with error bands ( $\pm 1$  s.d.). (C) MNIST  $N_I$  dose-response detail with error bands. (D) Fashion-MNIST individual seeds for all conditions, showing consistency across runs.

## A Supplementary Results

### Rule-Family Ranking

### Broadcast Mode Comparison and Local-Mismatch Recheck

### Phase 1 Capacity Ceilings

### Ablation Results

### Extended Gradient Analysis and $N_I$ Sweep Detail

### Verification and Reproducibility

### FA/DFA Baseline Comparison

### CIFAR-10 Results

### Additive Normalization Control

## B Implementation Details

### Biological Plausibility Assumptions

Each synapse has access to: presynaptic activity  $x_j$  (local), compartment voltage  $V_n$  (local membrane

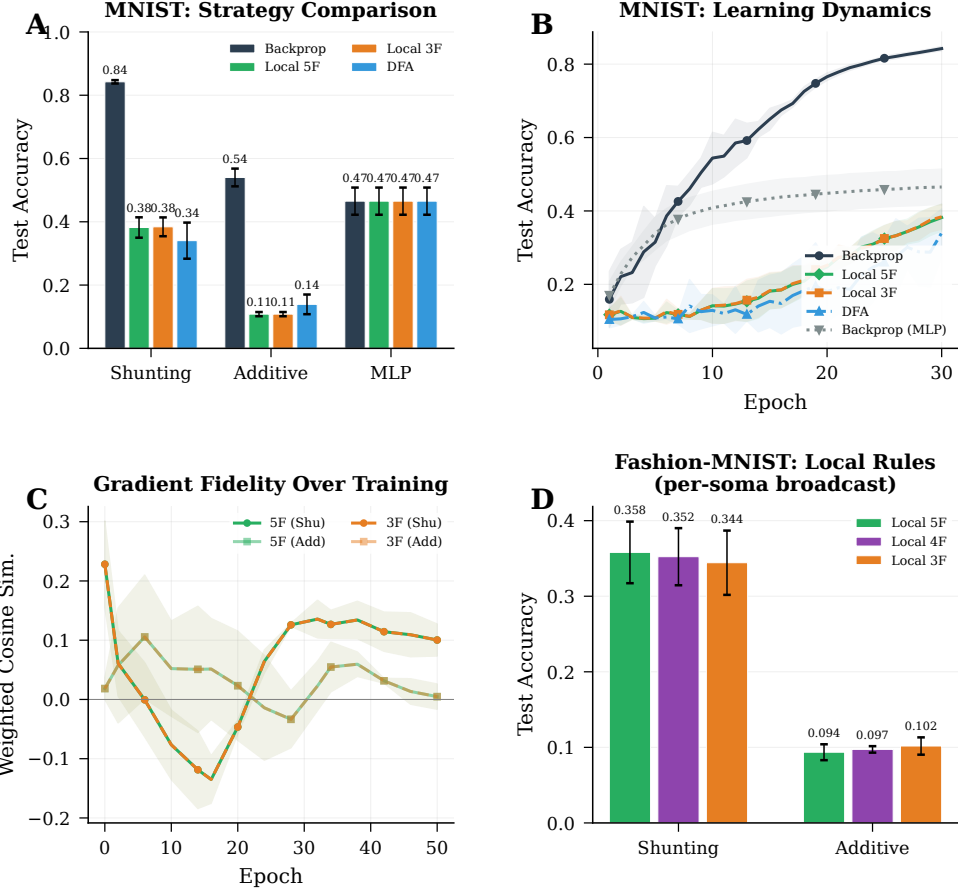


Figure S3: **Controlled small-network sandbox.** Strategy comparisons, learning dynamics, and gradient-fidelity trends.

Quantity	Symbol	Convention
Voltage	$V$	Normalized to $[-1, 1]$
Conductances	$g^{\text{syn}}, g^{\text{den}}$	Nonneg. via softplus
Leak conductance	$g^{\text{leak}}$	Set to 1
Input resistance	$R^{\text{tot}}$	$\leq 1$

Table S4: Units and normalization.

(computable from local conductances). The 4F modulator  $\rho_n$  requires online estimation of voltage covariance (biologically plausible via slow calcium signals); the 5F factor  $\phi_n$  requires a linear regression proxy (implementable via eligibility traces). The *only non-local quantity* is the broadcast error  $e_n$ , which requires a top-down or neuromodulatory signal from the soma to dendritic compartments. We assume per-neuron resolution (one error per output neuron) for the per-soma broadcast mode.

#### Units and Parameterization

#### Decoder Update Modes

$W_{\text{dec}}$  maps  $V_L \rightarrow \hat{y}$ . Modes: **backprop** ( $\nabla_W L$  via autograd), **local** ( $\Delta W = \eta \langle \delta_0 V_L^T \rangle_B$ ), **frozen** ( $\Delta W = 0$ ).

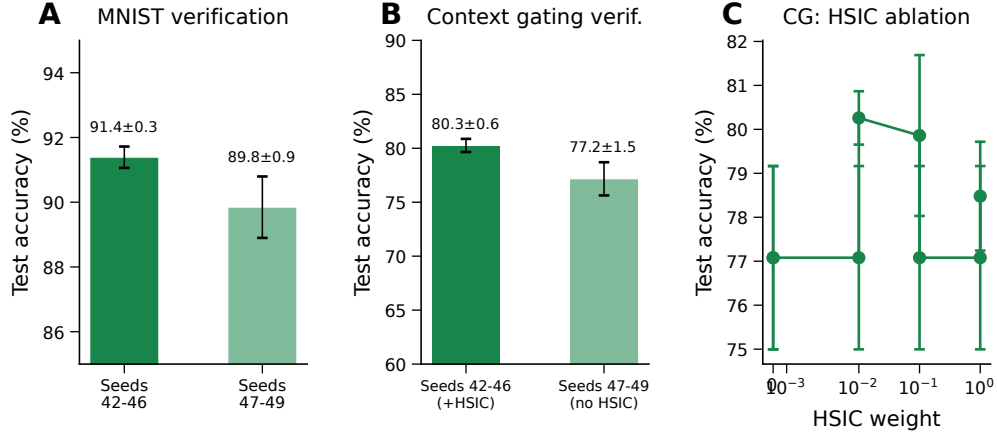


Figure S4: **Verification and reproducibility (supplementary).** (A) MNIST verification: main seeds (42–46) yield  $91.4 \pm 0.3\%$ ; held-out seeds (47–49) yield  $89.8 \pm 0.9\%$ , confirming generalization. (B) Context gating verification: main seeds (+HSIC)  $80.3 \pm 0.6\%$ ; held-out seeds (no HSIC)  $77.2 \pm 1.5\%$ . The  $\sim 3$  pp gap reflects HSIC removal, not seed sensitivity. (C) HSIC weight ablation on context gating: moderate weights (0.01–0.1) perform best.

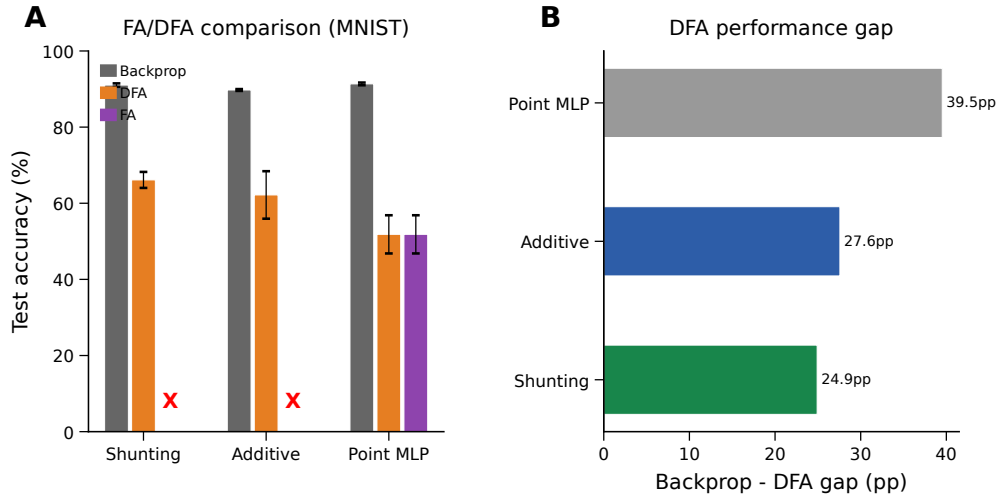


Figure S5: **Feedback alignment baselines (MNIST).** (A) Grouped comparison: standard backprop (gray), DFA (orange), and FA (purple). Red X marks indicate FA failure on dendritic architectures (dimensional incompatibility between random feedback matrices and block-structured dendritic layers). DFA achieves 66.1% on shunting vs. 62.2% additive vs. 51.8% point MLP. (B) Backprop–DFA gap: dendritic architectures (24.9–27.6 pp) show smaller gaps than point MLPs (39.5 pp), suggesting conductance-based architecture is partially compatible with random feedback.

## 348 Algorithm

### Algorithm 1 Local Credit Assignment

- 1: **Input:** Model, batch  $(x, y)$ , config  $\mathcal{C}$
- 2: Forward pass; loss  $L$ , output error  $\delta^y$
- 3: Somatic error  $\delta_0 = W_{\text{dec}}^\top \delta^y$
- 4: **for** each layer  $n$  (reverse) **do**
- 5:    $e_n = \text{broadcast}(\delta_0, \mathcal{C})$
- 6:   Compute  $\rho_n, \phi_n$  (EMA estimators)
- 7:   Apply 3F/4F/5F update (Eq. 6 or 7)
- 8: **end for**
- 9: Clip gradients; optimizer step

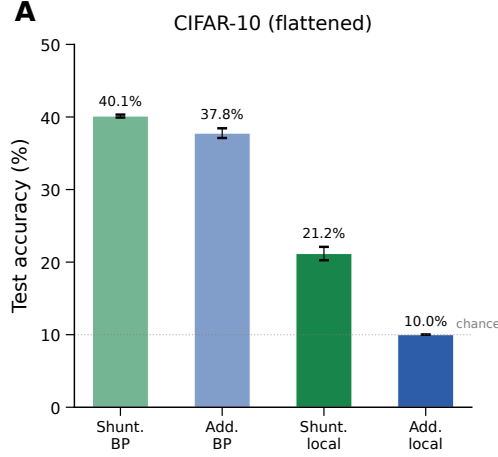


Figure S6: **CIFAR-10 (flattened)**. Shunting backprop: 40.1%; additive backprop: 37.8%; shunting local: 21.2%; additive local: 10.0% (chance). The backprop–local gap ( $\sim 19$  pp for shunting) is larger than on MNIST ( $\sim 0$  pp), reflecting the increased difficulty of propagating credit through conductance-based layers for complex visual features. Shunting still provides a clear advantage over additive under local learning ( $+11.2$  pp).

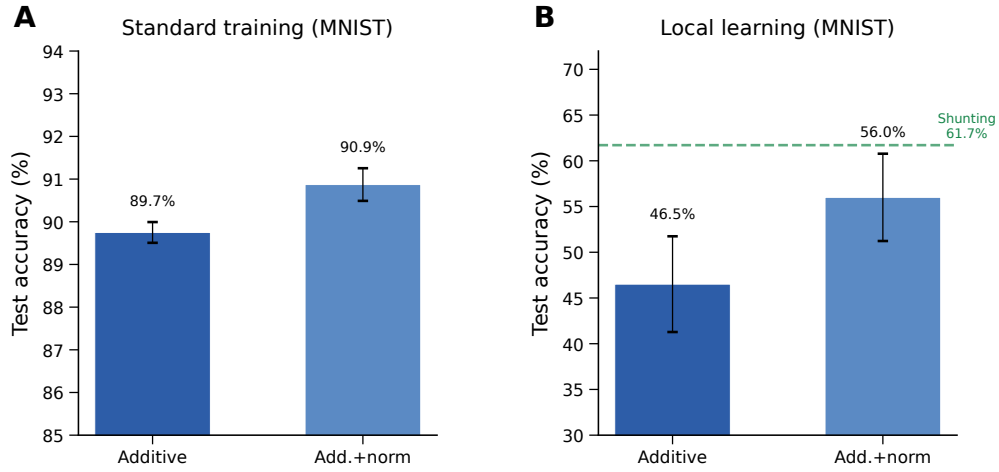


Figure S7: **Additive + normalization control (MNIST)**. (A) Standard backprop: normalization provides a small boost (89.7%  $\rightarrow$  90.9%). (B) Local learning: normalization improves additive from 46.5% to 56.0% ( $+9.5$  pp), partially closing the gap to shunting (61.7%, dashed green). The remaining gap ( $-5.7$  pp) indicates shunting provides benefits beyond simple voltage normalization, including bounded activations, conductance-dependent input weighting, and biophysically constrained sensitivity structure.

## 349 C Theoretical Details

### 350 Variant Taxonomy

### 351 Biological Analogs

## 352 D Morphology-Aware Extensions

353 **Path-integrated propagation.** Modulate broadcast error by  $\pi_n = \pi_{n-1} \cdot R_{n-1}^{\text{tot}} \cdot \bar{g}_{n-1}^{\text{den}}$ , approxi-  
 354 mating depth attenuation from Eq. 4.



Rule	Factors	Cost	Best regime
3F	$x, (E - V), e$	$\mathcal{O}(1)$	Baseline
4F	$3F + \rho$	$\mathcal{O}(1)$	Improved conditioning
5F	$4F + \phi$	$\mathcal{O}(d_n)$	Strongest overall

Table S5: Variant taxonomy.

Component	Analog	Interpretation
$R_n^{\text{tot}}$	Input resistance	Sensitivity modulation
$(E_j - V_n)$	Synaptic driving force	Local gradient factor
Shunting	Divisive normalization	$\partial V / \partial g_I \propto -V$
$\rho_n$	Layer relevance	Output correlation
$\phi_n$	Signal propagation	Conditional predictability

Table S6: Biological analogs.

355 **Depth modulation.** Per-branch scaling  $\rho_j = \rho_{\text{base}} / (d_j + \alpha)$ , mirroring cable attenuation.

356 **Dendritic normalization.**  $\Delta g_j^{\text{den}} \leftarrow \Delta g_j^{\text{den}} / (\sum_k g_k^{\text{den}} + \varepsilon)$ , analogous to homeostatic scaling  
357 [22].

358 **Apical/basal differentiation.** Branch-type scaling  $s_j$  for differential plasticity [20].

## 359 E HSIC Auxiliary Objectives

360 Following [17], we use kernel-based HSIC objectives. Self-decorrelation:  $\mathcal{L}^{\text{self}} =$   
361  $B^{-2} \text{tr}(\mathbf{K}_Z \mathbf{H} \mathbf{K}_Z \mathbf{H})$ . Target-correlation:  $\mathcal{L}^{\text{target}} = -B^{-2} \text{tr}(\mathbf{K}_Z \mathbf{H} \mathbf{K}_Y \mathbf{H})$ . Moderate weights  
362 (0.01–0.1) help on context gating; negligible on MNIST. Online statistics ( $\rho_n, \phi_n$ ) use Welford’s  
363 algorithm [21].

## 364 F Online Variant with Eligibility Traces

365 Continuous-time eligibility:  $\tau_e \dot{e}_j^{\text{syn}} = -e_j^{\text{syn}} + x_j(E_j - V_n)R_n^{\text{tot}}$ . Update:  $\Delta g_j^{\text{syn}} \propto$   
366  $\int e_j^{\text{syn}}(t)m_n(t) dt$  [18, 19].

## 367 G Depth Scaling and Noise Robustness

368 **Depth scaling.** Varying dendritic depth from 1–4 layers (branch factors [9] to [3, 3, 3, 3]): shunting  
369 local degrades from 63.5% to 57.4%; additive local degrades more steeply from 54.9% to 29.7%.  
370 The shunting advantage grows with depth (+8.5  $\rightarrow$  +27.7 pp). Backprop ceilings remain stable at  
371  $\sim 90$ – $92\%$ . See Fig. 4A.

372 **Noise robustness.** Gaussian noise  $\mathcal{N}(0, \sigma^2)$  on broadcast error: shunting is robust to  $\sigma \leq 0.1$   
373 ( $\sim 62\%$ ) while additive drops from 46.5% to chance at  $\sigma=1.0$ , confirming that shunting credit signals  
374 carry genuine learning information beyond the broadcast magnitude. See Fig. 4B.