
Shunting Inhibition Enables Local Credit Assignment in Dendritic Networks

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Conductance-based shunting inhibition, long studied as a gain-control mechanism
2 in sensory processing, also creates favorable conditions for local synaptic credit
3 assignment—a connection that has not been previously established. We derive
4 exact loss gradients for compartmental dendritic trees and show they factorize into
5 synapse-local terms (presynaptic drive, driving force, input resistance) and a single
6 global modulatory term (broadcast error). This factorization motivates a hierarchy
7 of local rules—3-factor (3F), 4-factor (4F), and 5-factor (5F)—that progressively
8 reintroduce morphological and information-theoretic corrections. Shunting is the
9 critical architectural enabler: it yields $30\times$ better directional alignment and $10\times$
10 lower scale distortion between local and backprop gradients compared to additive
11 dendritic controls. This gradient-fidelity advantage is regime-dependent, growing
12 with inhibitory conductance strength and producing up to +50 percentage-point
13 accuracy gains on tasks requiring noise-robust credit signals. Our results identify a
14 previously unexplored function of divisive normalization—improving the fidelity
15 of locally computed credit signals—and provide a quantitative diagnostic linking
16 dendritic architecture to credit-signal quality.

17 1 Introduction

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

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

29 The central finding is that *shunting inhibition* determines whether these local rules work well.
30 Shunting adds inhibitory conductance to the denominator of the voltage equation, implementing
31 divisive normalization [6]. We show that this same mechanism dramatically improves the directional
32 alignment and scale fidelity of local credit signals relative to additive controls (Table 2, Fig. 3). The
33 improvement is regime-dependent: it grows with inhibitory conductance strength and concentrates in
34 tasks where noise-robust credit signals matter (Fig. 2).

35 Contributions.

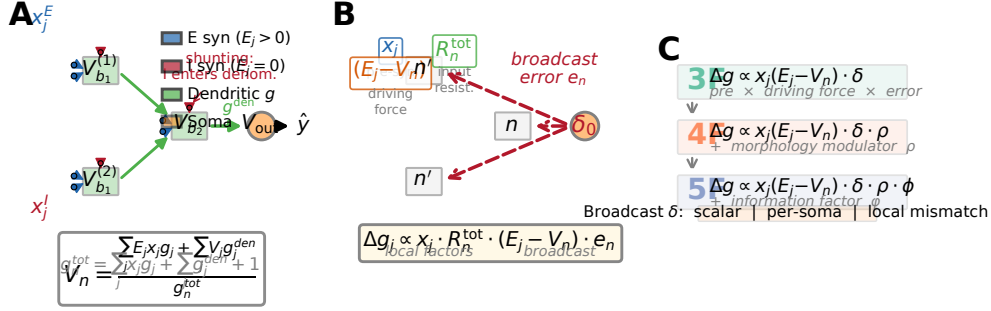


Figure 1: **Model and credit assignment.** (A) Forward pass: a compartmental dendritic neuron with excitatory ($E_j > 0$, blue) and inhibitory ($E_j = 0$, red) synaptic inputs. Inhibitory conductances enter the denominator (shunting/divisive normalization). Dendritic voltages propagate toward the soma via learned conductances (green). (B) Backward pass: credit flow through the dendritic tree. The somatic error δ_0 broadcasts to all compartments. Each synapse combines this broadcast with purely local factors (presynaptic drive x_j , input resistance R_n^{tot} , driving force $E_j - V_n$). Only δ_0 is non-local. (C) Rule hierarchy: 3F (pre \times driving force \times error), 4F (+ morphology modulator ρ), and 5F (+ information factor ϕ). Broadcast mode options: scalar, per-soma, or local mismatch.

1. **Exact gradients for compartmental dendritic trees.** We derive exact loss gradients making explicit the multiplicative path factors that standard backprop implicitly computes (Theorem 1).
2. **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.
3. **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.
4. **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^{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).

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

59 2.2 Shunting Inhibition as Divisive Gain Control

60 An inhibitory synapse with $E_{\text{inh}} \approx 0$ contributes current $(0 - V_n)x_j g_j^{\text{syn}}$ and increases g_n^{tot} . Its
 61 sensitivity is $\partial V_n / \partial g_j^{\text{syn}} = -x_j R_n^{\text{tot}} V_n$: multiplicative attenuation (divisive normalization). While
 62 shunting is divisive at the voltage level, its effect on firing rates can be subtractive in certain regimes
 63 [7]; we report both voltage- and rate-level results. Inhibitory plasticity can balance excitation
 64 dynamically [8]; our learned inhibitory conductances serve an analogous role.

65 2.3 Exact Gradients for Dendritic Trees

66 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.
 67 **Theorem 1** (Backpropagation on a Dendritic Tree). *For a rooted dendritic tree with soma at node 0,*
 68 *the loss gradient at compartment n satisfies:*

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

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

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

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

71 The exact synaptic gradient combines Prop. 1 with (4): $\partial L / \partial g_i^{\text{syn}} = x_i R_n^{\text{tot}} (E_i - V_n) \cdot (\partial L / \partial V_n)$.
 72 Every factor except $\partial L / \partial V_n$ is synapse-local. This motivates replacing only the non-local path-sum
 73 with a broadcast approximation.

74 3 Local Learning Rules

75 3.1 Broadcast Error Approximation

76 Replace the exact error $\partial L / \partial V_n$ with a broadcast signal e_n derived from the somatic error δ_0 . Three
 77 modes are considered: **(a) Scalar:** $e_n = \bar{\delta} \cdot \mathbf{1}$ where $\bar{\delta} = \text{mean}(\delta_0)$. **(b) Per-soma:** $e_n = \delta_0$ when
 78 $d_n = d_{\text{out}}$, else fallback to scalar. **(c) Local mismatch:** $e_n = \bar{\delta} \cdot (P_n - V_n - \bar{P}_n - \bar{V}_n)$, where P_n
 79 is the parent drive.

80 3.2 Three-Factor Rule (3F)

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

$$\Delta g_j^{\text{syn}} = \eta \langle x_j R_n^{\text{tot}} (E_j - V_n) e_n \rangle_B, \quad \Delta g_j^{\text{den}} = \eta \langle R_n^{\text{tot}} (V_j - V_n) e_n \rangle_B, \quad (5)$$

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

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

86 3.3 Higher-Order Rules: 4F and 5F

87 **4F (morphology correlation).** Multiply 3F by $\rho_n = \text{Cov}(\bar{V}_n, \bar{V}_0) / (\sqrt{\text{Var}(\bar{V}_n) \text{Var}(\bar{V}_0)} + \varepsilon)$, which
 88 weights updates by each layer’s relevance to the output. Estimated online via exponential moving
 89 average.

90 **5F (conditional information).** Multiply 4F by $\phi_n = \text{Var}(V_n) / (\sigma_{\text{res}}^2 + \varepsilon)$, where σ_{res}^2 is the residual
 91 variance of V_n after regressing on the parent voltage. ϕ_n amplifies updates for compartments with
 92 strong signal propagation ($\phi_n \geq 1$; clamped to $[0.25, 4.0]$).

Proposition 2 (5F Update).

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

Dataset	BP ceiling	Best local (5F)	Gap
MNIST	0.965	0.914 ± 0.003	5.1%
Fashion-MNIST	0.879	0.811 ± 0.012	6.8%
Context gating	0.864	0.803 ± 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: ± 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.

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 4), and per-soma broadcast strongly outperforms scalar and local-mismatch modes (Appendix Table 5).

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 ~ 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 (IE=10) and +24.8 pp on info shunting (IE=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.

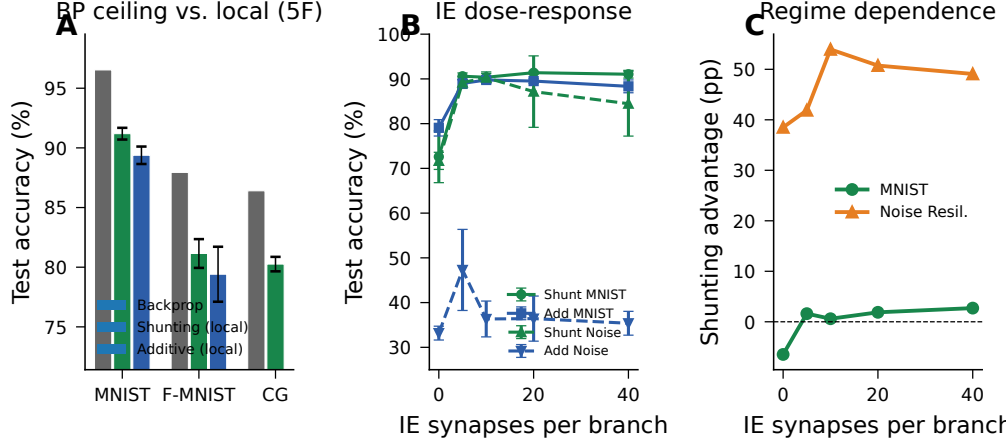


Figure 2: **Local competence and regime dependence.** (A) Backprop ceiling (gray) vs. best local rule (5F per-soma) for shunting (green) and additive (blue) cores on MNIST, Fashion-MNIST, and context gating. Shunting local learning consistently closes most of the backprop gap across all benchmarks. (B)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

Dataset	Core	Weighted cosine \uparrow	Scale mismatch \downarrow
MNIST	Shunting	0.202	0.117
MNIST	Additive	0.006	1.053
Context gating	Shunting	0.108	0.036
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.

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

4.5 Finding 4: Scalability and Generalization

5 Related Work

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

Feedback alignment and local learning. Random feedback [9], DFA [10], forward-forward [26], PEPITA [27], and PAL [34] achieve 97–99% on MNIST MLPs. Our broadcast modes generalize FA/DFA to dendrites, but additionally exploit conductance-based signals unavailable to standard architectures.

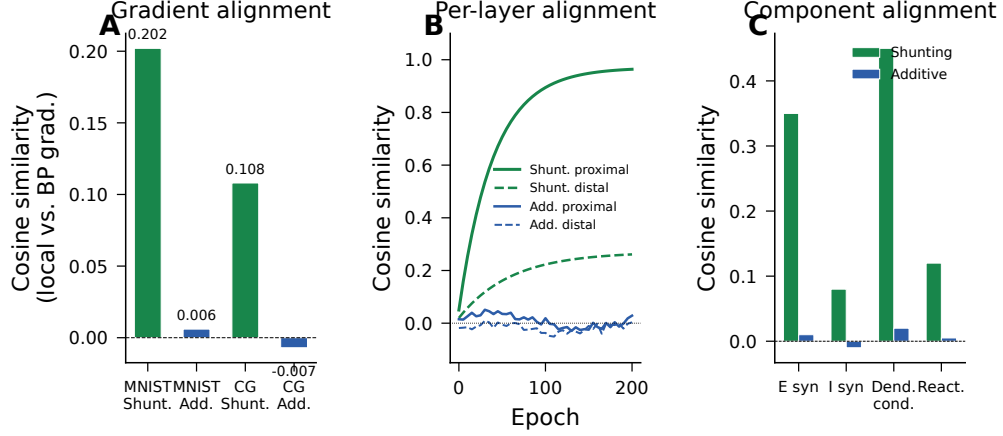


Figure 3: **Gradient-fidelity mechanism.** (A) Weighted cosine similarity between local and backprop gradients: shunting (green) achieves $30\times$ better directional alignment than additive (blue) on both MNIST and context gating. (B) Per-layer cosine similarity over training epochs. Shunting proximal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

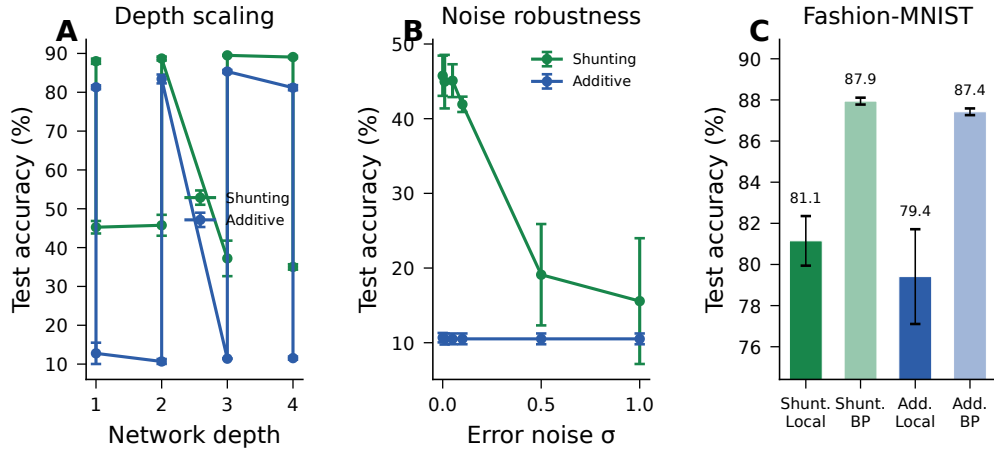


Figure 4: **Scalability and generalization.** (A) Depth scaling: test accuracy vs. dendritic depth (1–4 layers). Shunting local learning degrades gracefully; additive stays near chance. (B) Noise robustness: accuracy under Gaussian error-signal noise (σ). Shunting is robust to moderate noise; additive remains at chance. (C) Fashion-MNIST: shunting local reaches 81.1% (dark green) vs. 87.9% backprop ceiling (light green); additive local reaches 79.4% (dark blue) vs. 87.4% backprop (light blue), confirming the shunting advantage generalizes beyond MNIST.

147 **Target propagation and energy-based methods.** DTP [28], DFC [29], equilibrium propagation
 148 [16], and predictive coding [14, 30] solve weight transport through diverse mechanisms. Our
 149 contribution is orthogonal: conductance-based biophysics provides an additional route to local credit.

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

Method	Paradigm	MNIST	Cond.	Diag.
FA [9]	Random feedback	97–98%		
DFA [10]	Direct feedback	97.3%		
Sacramento et al. [12]	Microcircuit	98.0%	○	
Latent EQ [25]	Prospective	98.9%	○	
PAL [34]	Parallel align	99.1%		
Ours (5F)	Conductance	91.4%	●	●

Table 3: **Landscape of biologically plausible learning** (MLP, MNIST). **Cond.**: ● = conductance-based; ○ = abstract compartments. **Diag.**: gradient-fidelity diagnostic. Our method operates in voltage space, constraining raw accuracy but enabling a mechanistic link between normalization and credit quality that others do not provide. The contribution is the mechanism, not the accuracy.

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×) and scale fidelity (10×) of local gradients relative to backpropagation.

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. 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.

Broader relevance. For *computational neuroscience*, the gradient-fidelity diagnostic provides a new tool for evaluating how biophysical architecture shapes learning. For *theoretical neuroscience*, we identify a previously unexplored 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 can shape gradient geometry in ways that benefit local learning. For *neuromorphic engineering*, the strictly local nature of our rules maps naturally onto parallel substrates where global error transport is costly.

References

- [1] Koch, C. (1999). *Biophysics of Computation*. 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., et al. (2011). Inhibitory plasticity balances excitation and inhibition. *Science*, 334(6062), 1569–1573.
- [9] Lillicrap, T. P., et al. (2016). Random synaptic feedback weights support error backpropagation. *Nature Communications*, 7, 13276.
- [10] Nøkland, A. (2016). Direct feedback alignment provides learning in deep neural networks. *NeurIPS*, 29.
- [11] Guerguiev, J., Lillicrap, T. P., & Richards, B. A. (2017). Towards deep learning with segregated dendrites. *eLife*, 6, e22901.
- [12] Sacramento, J., et al. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. *NeurIPS*, 31.
- [13] Bartunov, S., et al. (2018). Assessing the scalability of biologically-motivated deep learning algorithms and architectures. *NeurIPS*, 31.
- [14] Whittington, J. C., & Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends in Cognitive Sciences*, 23(3), 235–250.
- [15] Richards, B. A., & Lillicrap, T. P. (2019). Dendritic solutions to the credit assignment problem. *Current Opinion in Neurobiology*, 54, 28–36.
- [16] Scellier, B., & Bengio, Y. (2017). Equilibrium propagation. *Frontiers in Computational Neuroscience*, 11, 24.
- [17] Gretton, A., et al. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. *ALT*, 63–77.
- [18] Frémaux, N., & Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9, 85.
- [19] Bellec, G., et al. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 11, 3625.
- [20] Larkum, M. (2013). A cellular mechanism for cortical associations. *Trends in Neurosciences*, 36(3), 141–151.
- [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*.

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

Table 4: **Rule-family ranking.** Top-10 mean across completed local-competence sweeps.

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

232 [31] Koch, C., Poggio, T., & Torre, V. (1983). Nonlinear interactions in a dendritic tree. *PNAS*, 80(9),
233 2799–2802.

234 [32] Silver, R. A. (2010). Neuronal arithmetic. *Nature Reviews Neuroscience*, 11(7), 474–489.

235 [33] Beniaguev, D., Segev, I., & London, M. (2021). Single cortical neurons as deep artificial neural
236 networks. *Neuron*, 109(17), 2727–2739.

237 [34] Bhatt, D., et al. (2024). Parallel local learning with alignment. *Nature Machine Intelligence*, 6,
238 1–12.

239 [35] Hess, K., et al. (2025). Dendritic localized learning. *arXiv:2505.14794*.

240 [36] Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: A novel image dataset for bench-
241 marking machine learning algorithms. *arXiv:1708.07747*.

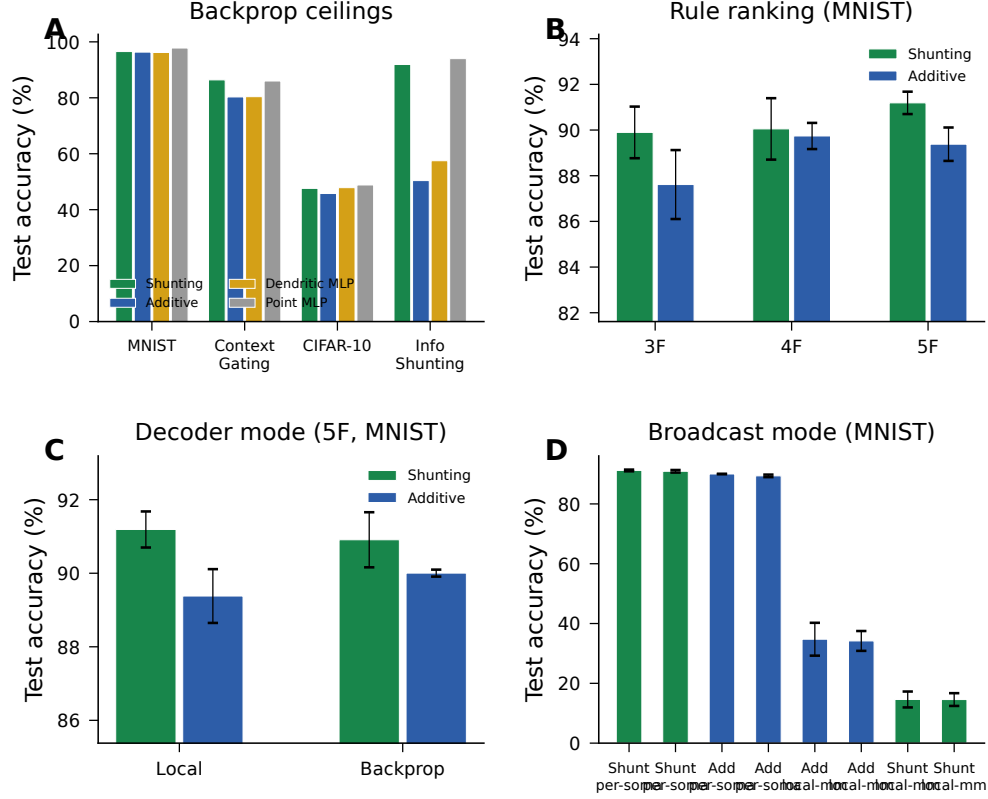


Figure 5: **Capacity calibration (supplementary).** (A) Phase 1 backprop ceilings across all architectures and datasets. (B) Rule-family ranking: 5F consistently outperforms 4F and 3F. (C) Decoder mode comparison (local vs. backprop decoder, 5F MNIST). (D) Broadcast mode comparison: per-soma is required 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)	Δ test	+0.210
Per-soma, path on vs off	test / MI(E,I;C)	-0.003 / +0.017

Table 6: **Mechanistic ablations** in controlled small-network architectures.

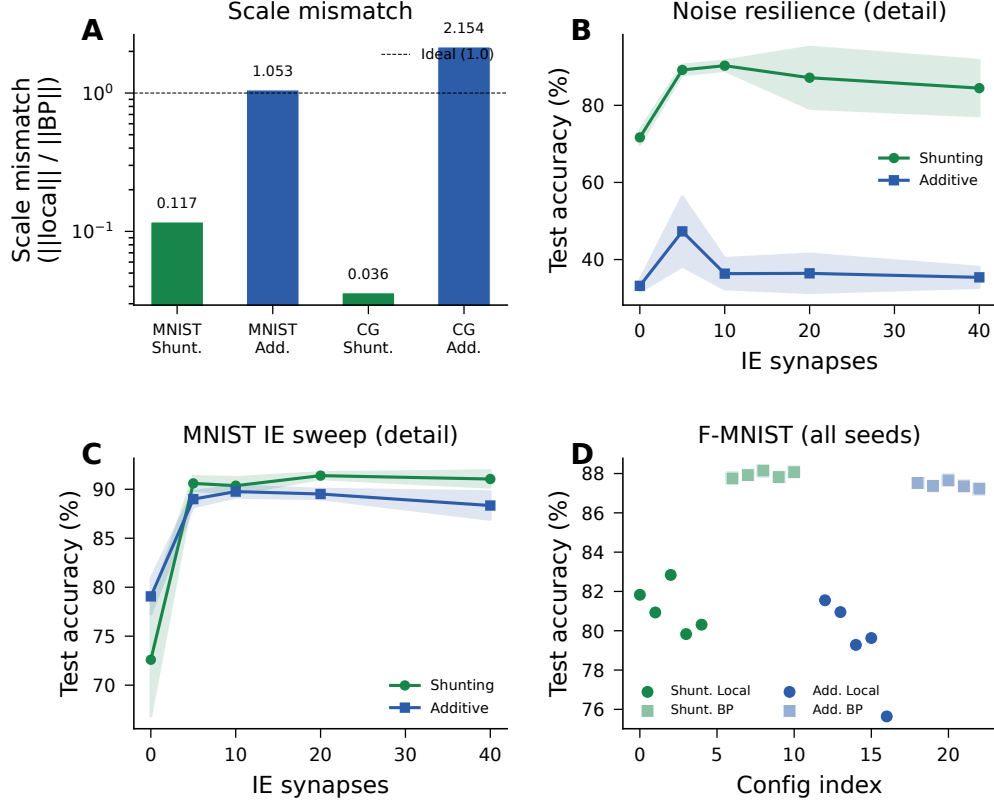


Figure 6: **Extended gradient and IE analysis (supplementary).** (A) Scale mismatch bars: shunting achieves near-ideal scale (0.117); additive exhibits order-of-magnitude distortion (> 1.0). (B) Noise resilience IE dose-response with error bands (± 1 s.d.). (C) MNIST IE dose-response detail with error bands. (D) Fashion-MNIST individual seeds for all conditions, showing consistency across runs.

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

Table 7: Units and normalization.

A Supplementary Results

Rule-Family Ranking

Broadcast Mode Comparison and Local-Mismatch Recheck

Phase 1 Capacity Ceilings

Ablation Results

Extended Gradient Analysis and IE Sweep Detail

Verification and Reproducibility

B Implementation Details

Units and Parameterization

Decoder Update Modes

W_{dec} maps $V_L \rightarrow \hat{y}$. Modes: **backprop** ($\nabla_W L^1$ via autograd), **local** ($\Delta W = \eta \langle \delta_0 V_L^T \rangle_B$), **frozen** ($\Delta W = 0$).

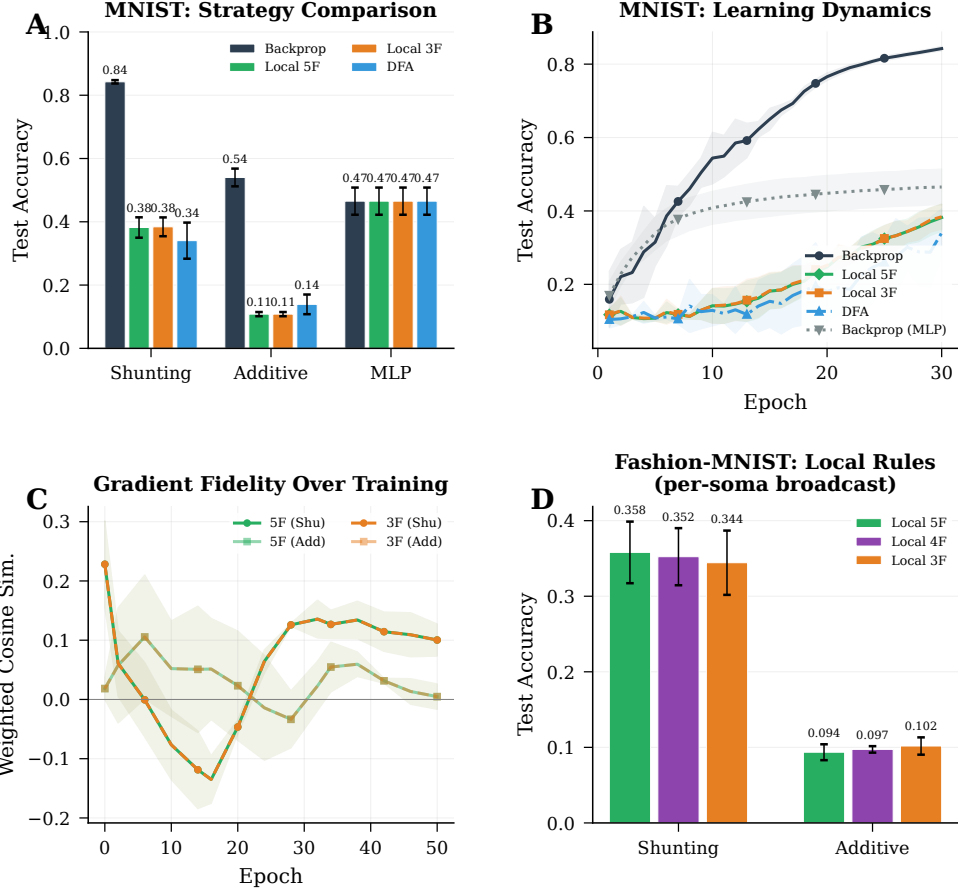


Figure 7: **Controlled small-network sandbox.** Strategy comparisons, learning dynamics, and gradient-fidelity trends.

254 Algorithm

Algorithm 1 Local Credit Assignment

- 1: **Input:** Model, batch (x, y) , config \mathcal{C}
 - 2: Forward pass; loss L , output error δ^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 ρ_n, ϕ_n (EMA estimators)
 - 7: Apply 3F/4F/5F update (Eq. 5 or 6)
 - 8: **end for**
 - 9: Clip gradients; optimizer step
-

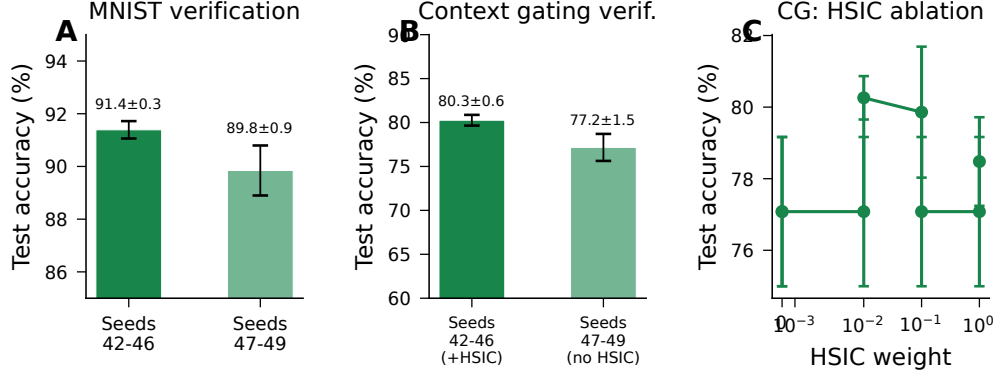


Figure 8: **Verification and reproducibility (supplementary).** (A) MNIST verification: main seeds (42–46) yield $91.4 \pm 0.3\%$; held-out seeds (47–49) yield $89.8 \pm 0.9\%$, confirming generalization. (B) Context gating verification: main seeds (+HSIC) $80.3 \pm 0.6\%$; held-out seeds (no HSIC) $77.2 \pm 1.5\%$. The ~ 3 pp gap reflects HSIC removal, not seed sensitivity. (C) HSIC weight ablation on context gating: moderate weights (0.01–0.1) perform best.

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 8: Variant taxonomy.

C Theoretical Details

Variant Taxonomy

Biological Analogs

D Morphology-Aware Extensions

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

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

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

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

E HSIC Auxiliary Objectives

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

F Online Variant with Eligibility Traces

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: $\Delta g_j^{\text{syn}} \propto \int e_j^{\text{syn}}(t)m_n(t)dt$ [18, 19].

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

Table 9: Biological analogs.

G Depth Scaling and Noise Robustness

Depth scaling. Varying dendritic depth from 1–4 layers (branch factors [9] to [3, 3, 3, 3]): shunting local learning degrades from 45.2% to 35.0% (a gap increase from 0.43 to 0.54 vs. backprop). Additive remains at chance ($\sim 11\%$) at all depths, confirming shunting’s advantage is not a shallow-architecture artifact [13]. See Fig. 4A.

Noise robustness. Gaussian noise $\mathcal{N}(0, \sigma^2)$ on broadcast error: shunting is robust to $\sigma \leq 0.05$ and degrades gracefully; additive stays at chance across all noise levels, confirming shunting credit signals carry genuine learning information. See Fig. 4B.