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 ± 0.003 test accuracy on MNIST and 0.803 ± 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.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

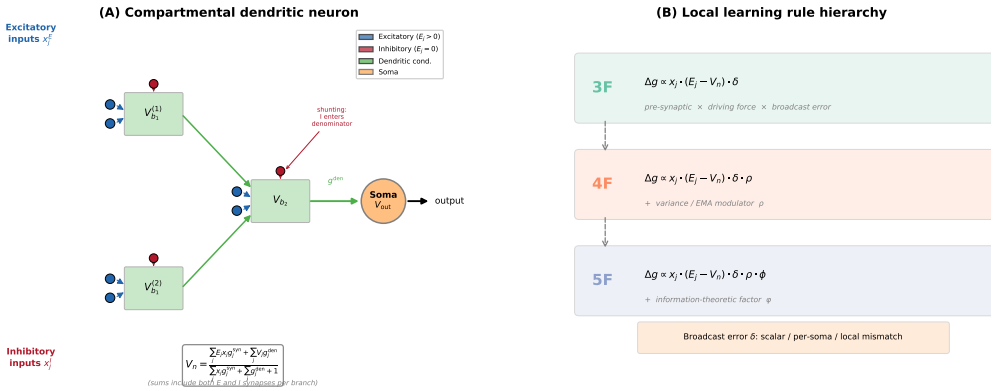


Figure 1: **Model overview.** (A) A compartmental dendritic neuron where *each branch* receives both excitatory ($E_j > 0$, blue) and inhibitory ($E_j = 0$, red) synaptic inputs via separate sparse connectivity (TopK). Inhibitory conductances enter only the denominator of the voltage equation (shunting/divisive 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 ρ), and 5-factor (+ information-theoretic factor ϕ). 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 δ can operate in scalar, per-soma, or local mismatch mode.

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

Three empirical findings (this draft).

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

2 Compartmental Voltage Model

We use a standard steady-state conductance model obtained by discretizing passive cable dynamics (e.g., [1, 2]). In normalized units with leak reversal potential 0 and unit leak conductance, each compartment voltage is a conductance-weighted average of synaptic reversal potentials, child voltages, and leak. This form makes 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 $E_{\text{inh}} \approx 0$ (Section 2.3).

2.1 Voltage Equation

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

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

Currents. Synaptic current:

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

Dendritic current:

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

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

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

Total conductance and resistance.

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

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

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

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

2.2 Local Sensitivities

Proposition 1 (Synaptic Gradient)**.**

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

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

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

Proposition 3 (Dendritic Gradient)**.**

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

2.3 Shunting Inhibition and Divisive Gain Control

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

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

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

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

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

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

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

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

2.4 Loss Propagation

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

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

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

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

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

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

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

3 Local Learning Approximations

3.1 Broadcast Error Approximation

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

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

Three broadcast modes are considered:

(A) Scalar broadcast. For minibatch index b:

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

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

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

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

Then:

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

3.2 Gradient Alignment with Broadcast Errors

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

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

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

where c_n depends on α and the average correlation between $\widehat{\Xi}_n$ and Ξ_n. Thus g^{local} provides a descent direction in expectation.

Sketch. Adapt the feedback-alignment argument [9, 10]: fixed random feedback suffices for alignment as forward weights adapt. Here, $\widehat{\Xi}_n$ is proportional to Ξ_n up to the missing path factor; Jensen bounds on (10) yield $c_n > 0$. □

3.3 Three-Factor Rule (3F)

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

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

For dendritic conductances:

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

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

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

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

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

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 in layer n. Define the correlation with output:*

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

Proposition 5 (4F Update Rule)**.** *Multiply 3F updates by ρ_n:*

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

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

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

Estimators (EMA / online). For minibatches $B \geq 2$, estimate ρ_n from (17) with an EMA over batches. For $B = 1$ (online), maintain means μ_x, μ_y, variances σ_x^2, σ_y^2, and covariance C_{xy} for $x_t = \bar{V}_0^{(t)}$ and $y_t = \bar{V}_n^{(t)}$ using Welford's numerically stable algorithm [20]:

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

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

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

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

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

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

The information proxy is:

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

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

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

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

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

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

4 Morphology-Aware Extensions

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

4.1 Path-Integrated Propagation

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

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

where $\bar{g}_{n-1}^{\text{den}}$ is the mean dendritic conductance from layer $n-1$ to n. In practice, π_n is computed per 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 ± 0.003	-0.051
Context gating	0.864	0.803 ± 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.

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*

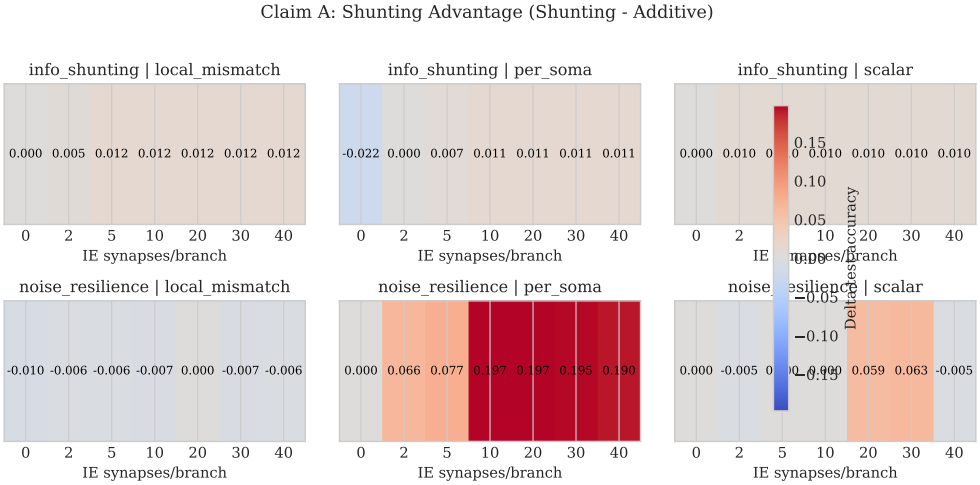


Figure 2: **Shunting advantage across inhibition strength.** Heatmap of shunting minus additive test-accuracy difference over inhibitory synapse count and broadcast modes, showing the regimes where shunting-linked local credit is most beneficial.

Dataset	Core	Weighted cosine	Scale mismatch	Rel. L2
MNIST	Shunting	**0.202**	**0.117**	**1.130**
MNIST	Additive	0.006	1.053	13.324
Context gating	Shunting	**0.108**	**0.036**	**1.404**
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.

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

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

and a scale mismatch

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

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

Limitations and future work. We view these results as evidence for a mechanism (shunting-linked credit stabilization) rather than a complete biological account. Remaining extensions include scaling to deeper architectures and vision benchmarks beyond flattened inputs, testing on reconstructed morphologies, and connecting the discrete-time rules to event-driven spiking implementations (Appendix F).

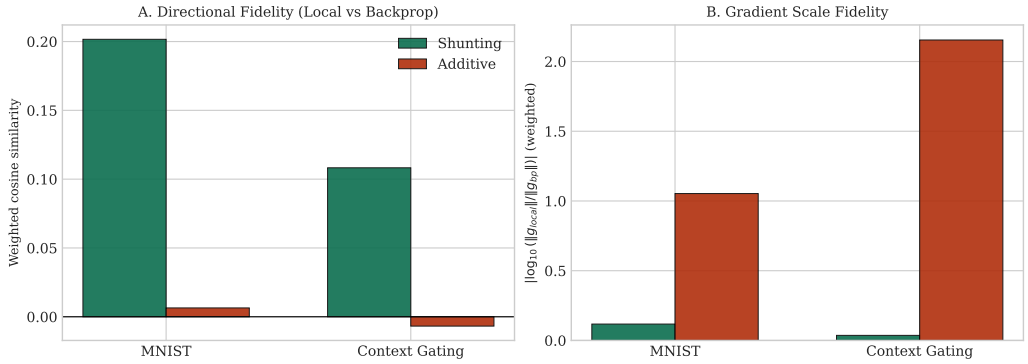


Figure 3: **Phase-2b best-regime gradient fidelity.** Shunting models show substantially better local-vs-backprop directional alignment (higher weighted cosine) and much lower gradient-scale distortion than additive controls on both MNIST and context-gating best configurations.

6 Related Work

Dendritic credit assignment. Dendritic trees support nonlinear subunit computation and can be interpreted as a multi-stage function approximator [3, 4]. Several proposals exploit dendritic compartmentalization for biologically plausible learning. Urbanczik & Senn [5] derive a single-neuron learning 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, enabling approximate deep learning. Sacramento et al. [12] show that cortical microcircuits with dendritic predictions can approximate backprop if apical and basal inputs converge with appropriate timing. Our work differs in two key ways: (i) we derive exact gradients for a conductance-based compartment model (not an abstract compartmental surrogate), and (ii) we identify shunting inhibition as a critical enabler of local gradient quality via quantitative gradient-fidelity analysis.

Feedback alignment and variants. Lillicrap et al. [9] showed that random fixed feedback weights suffice for learning in deep networks (feedback alignment, FA), and Nøkland [10] extended this to direct feedback alignment (DFA) where output errors project directly to each layer. Our broadcast error modes (Section 3.1) generalize this idea to the dendritic setting. The key difference is that our local rules also exploit conductance-based local signals (driving force, input resistance) that are unavailable to standard FA/DFA.

Other biologically plausible methods. Equilibrium propagation [14] computes exact gradients in energy-based networks by contrasting free and clamped phases. Predictive coding networks [13] perform inference via local prediction errors that converge to backprop gradients at equilibrium. Target propagation methods use local targets rather than error gradients. These approaches solve the weight transport problem through different mechanisms; our contribution is orthogonal, showing that dendritic biophysics provides yet another route to local credit assignment.

Divisive normalization and shunting. Shunting inhibition provides divisive normalization of neural responses [6], though its effect on firing rates can be subtractive in certain regimes [7]. Inhibitory plasticity and E/I balance are also implicated in stabilizing cortical dynamics [8]. We show that this divisive interaction has a benefit for learning: it creates regimes where local gradient approximations are substantially more faithful to exact gradients. Homeostatic mechanisms including synaptic scaling [21] and apical/basal plasticity differences [19] motivate our morphology-aware extensions.

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.

Dataset	Rule	Top-10 valid mean	Top-10 test mean
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 3: **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.

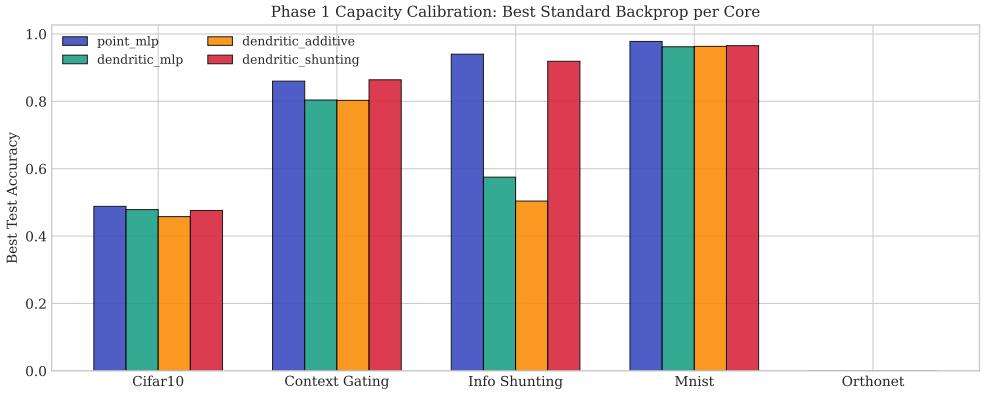


Figure 4: **Phase 1 capacity ceilings (standard backprop).** Best standard-test accuracy per dataset and core type.

[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.

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 strongly outperforms scalar and local-mismatch modes for both datasets (MNIST: 0.918 vs 0.889 vs 0.200 test; context gating: 0.827 vs 0.699 vs 0.114 test), motivating broadcast mode as a primary experimental factor.

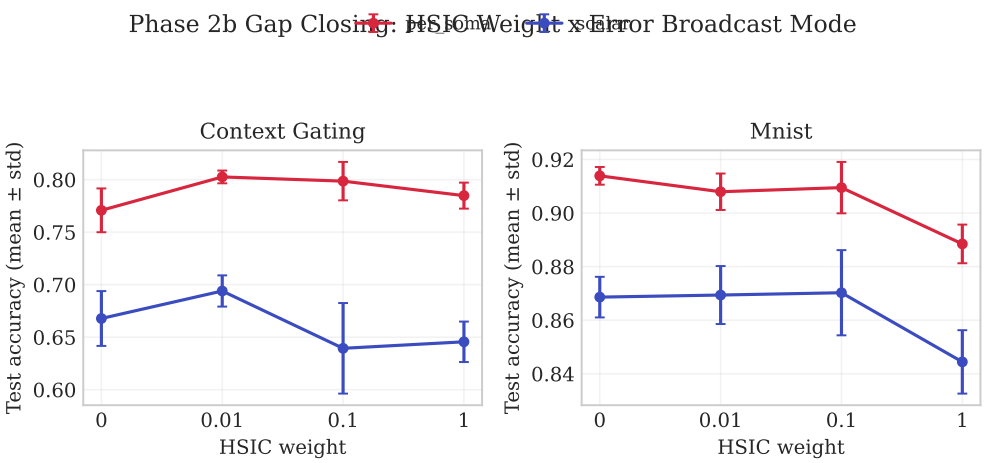


Figure 5: **HSIC strength and broadcast mode.** For context gating, moderate HSIC weights (0.01–0.1) improve local learning under per-soma broadcast, while large weights degrade performance. For MNIST, HSIC has negligible effect. Error bars: ± 1 s.d. across seeds.

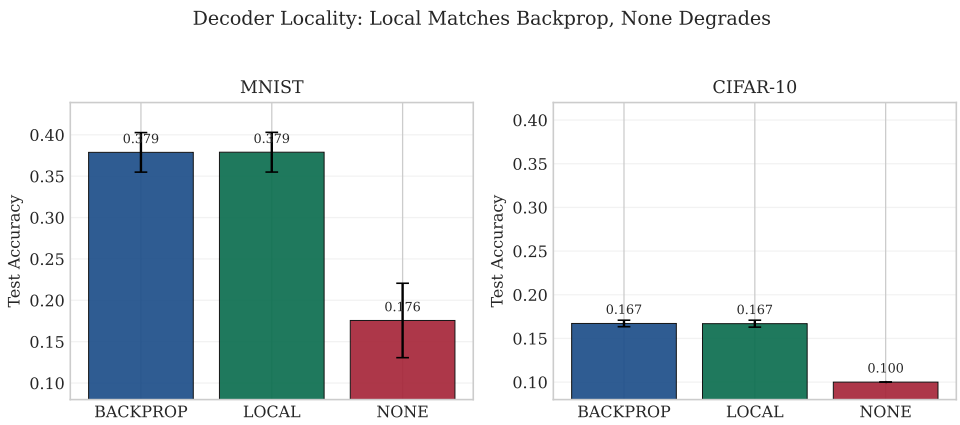


Figure 6: **Decoder locality.** On both MNIST and CIFAR-10, local decoder updates match backpropagated decoder updates, while frozen decoder weights collapse performance.

Phase-Based Sweep Figures

Additional Gradient-Fidelity Analyses

Small-Network Sandbox Summary

B Implementation Details (Appendix)

Units and Normalization

Positive Parameterization

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

Decoder Update Modes

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

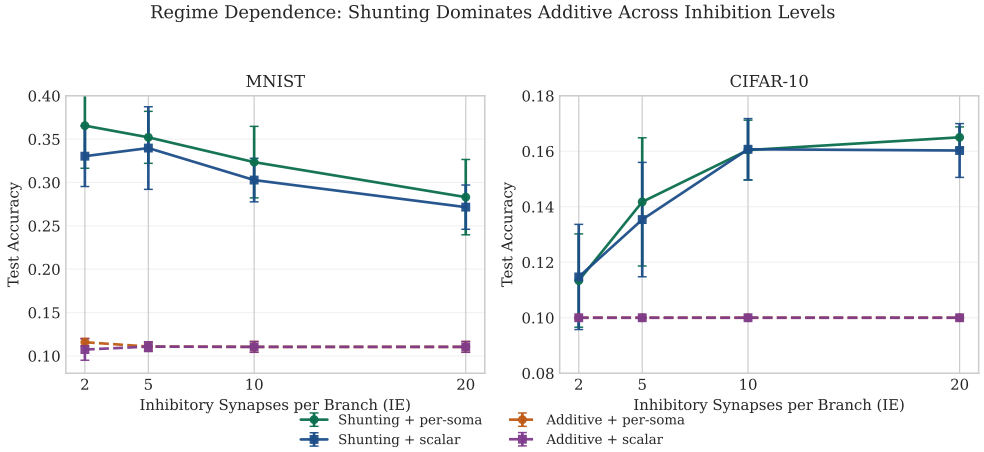


Figure 7: **Regime dependence across inhibition levels (robust sweep).** Shunting networks consistently outperform additive controls across inhibitory synapse counts and broadcast modes.

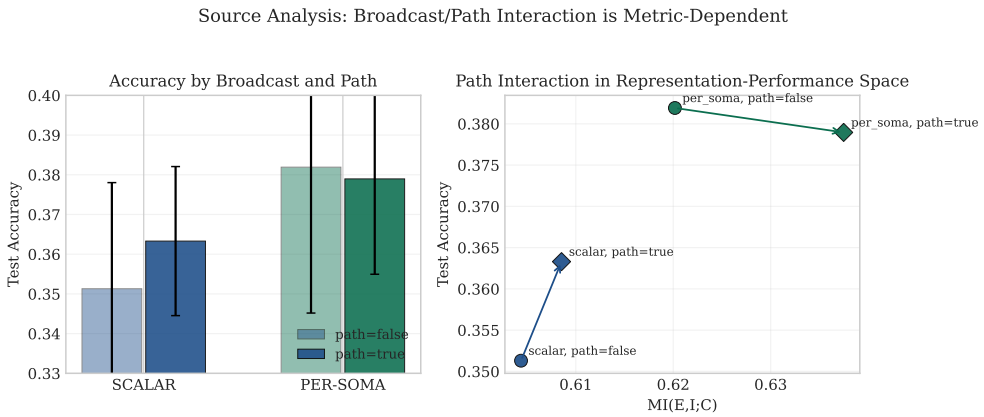


Figure 8: **Broadcast-path interaction in source analysis (robust sweep).** Within per-soma broadcast, path propagation changes information metrics more than accuracy; within scalar broadcast it improves accuracy more with smaller information gains.

Algorithm Sketch

Algorithm 1 Local Credit Assignment (schematic)

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

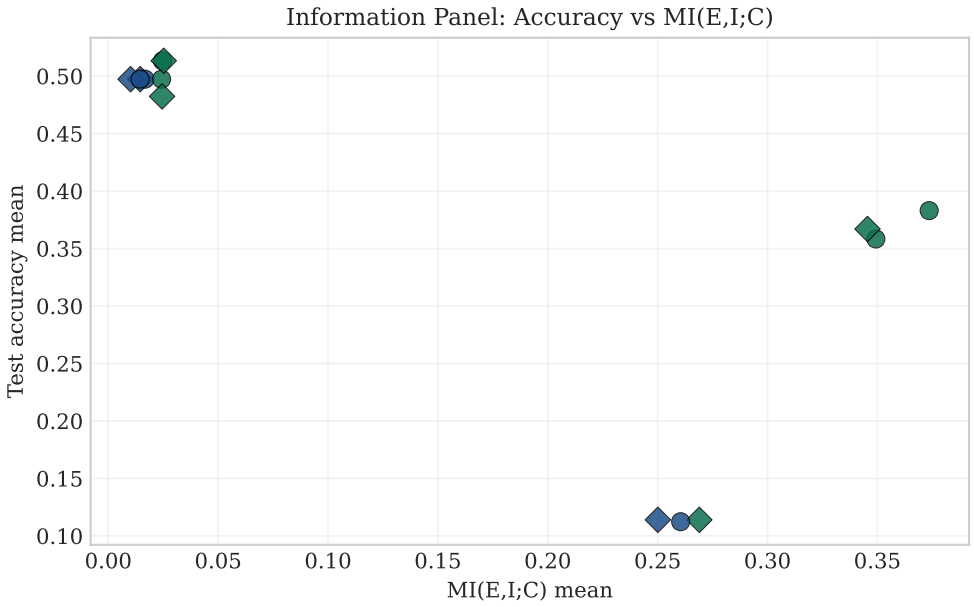
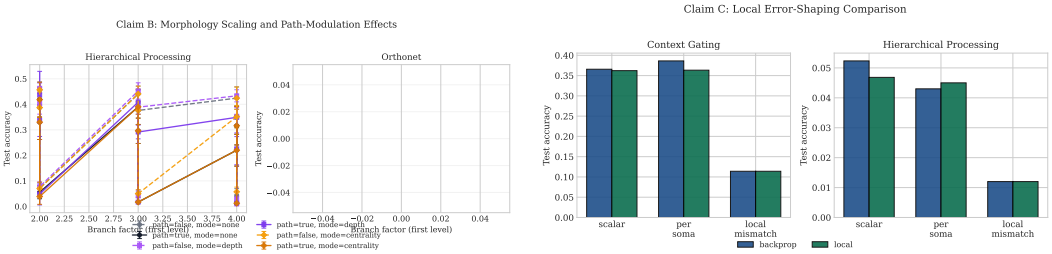


Figure 9: **Information panel.** Mutual information proxy $I(E, I; C)$ versus test accuracy for shunting (green) and additive (blue) networks, with and without path propagation.



(a) **Morphology scaling.** Accuracy as a function of dendritic branching with and without path propagation/modulators.

(b) **Error shaping.** Comparison of broadcast modes and decoder update modes on hierarchical/context tasks.

Figure 10: **Phase 3 ablations: morphology and error shaping.**

C Theoretical Comparison (Appendix)

D Morphology-Aware Extensions (Details)

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

Path-integrated propagation. Exact gradients in dendritic trees involve a path-sum of multiplicative edge factors (Eq. 10). Path propagation approximates this depth attenuation by modulating the broadcast error with the recursive path factor π_n (Eq. 26), implemented as a per-sample scalar broadcast within each layer.

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 $\|\Delta g_j^{\text{syn}}\| \propto 1/(d_j + \alpha)$.

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 / $\mathrm{MI}(E, I; C)$	-0.0030 / +0.0173

Table 4: **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.

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

Dendritic normalization. Normalize dendritic updates by total branch conductance: $\Delta g_j^{\mathrm{den}} \leftarrow \Delta g_j^{\mathrm{den}} / (\sum_k g_k^{\mathrm{den}} + \varepsilon)$. 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 $s_j = s_{\mathrm{basal}} + t_j(s_{\mathrm{apical}} - s_{\mathrm{basal}})$. Setting $s_{\mathrm{apical}} > s_{\mathrm{basal}}$ 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:

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

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

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

F Online Variant with Eligibility Traces

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

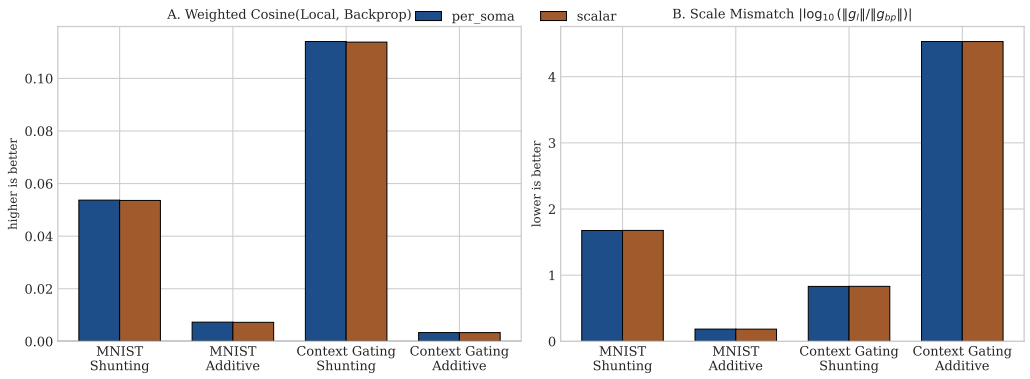


Figure 11: **Per-soma vs scalar gradient-fidelity on trained Phase-2b checkpoints.** Aggregate directional and scale metrics for shunting and additive cores on MNIST and context gating.

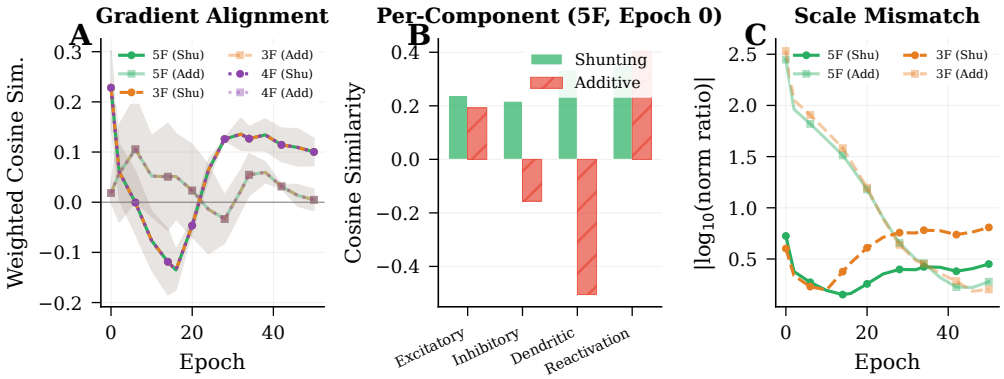


Figure 12: **Gradient fidelity over training.** (A) Parameter-count-weighted cosine similarity between local and backprop gradients over training epochs. (B) Per-component cosine at initialization. (C) Scale mismatch over training.

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

Table 6: 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, $\rho \to \rho_j$	$\mathcal{O}(d_n)$	Useful in deeper/branched morphologies
5F + Norm	5F + 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 7: Variant taxonomy: computational cost and current empirical regime map (L = depth, d_n = compartments).

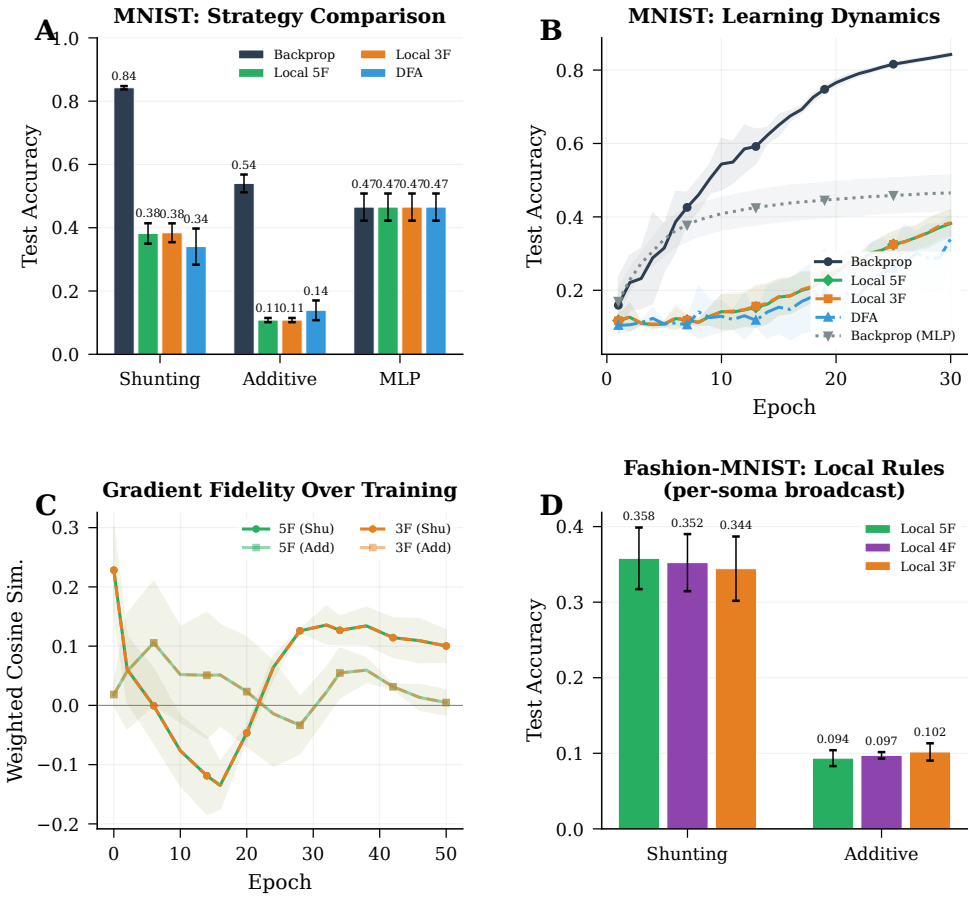


Figure 13: **Controlled small-network sandbox (summary).** A compact view of strategy comparisons, learning dynamics, and gradient-fidelity trends in a matched small-architecture regime used for mechanism isolation.

Component	**Biological Analog**	**Interpretation**
Conductance scaling R_n^{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 π_n	Cable attenuation	Approximate depth attenuation (Eq. 26)
Morphology factor ρ_n	Layer relevance	Correlation with output (Eq. 17)
Information factor ϕ_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 8: Theoretical components and their biological/algorithmic interpretations.