Morpho-Temporal Decoupling: How Primate Neurons Expand Dendrites Without Losing Speed

Jiawei Zhang[1,2,3†] Zhangjw22@m.fudan.edu.cn

Haoyu Albert Wang[1,3,†] haoyuwang18@fudan.edu.cn

Jialun Ma[1] 21307130025@m.fudan.edu.cn

Wei P. Dai[1,‡] weidai@fudan.edu.cn

Yuguo Yu[1,2,3,‡] yuyuguo@fudan.edu.cn

1. Research Institute of Intelligent Complex Systems, Fudan University.
2. State Key Laboratory of Brain Function and Disorders and MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University.
3. Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University.
† These authors contributed equally to this work.
‡ Corresponding author.

Abstract

During evolution, biological neurons scale computation by expanding dendritic branches. However, this expansion creates a biophysical dilemma: increased membrane area typically imposes a capacitive load that slows somatic dynamics and narrows temporal bandwidth. Here, we investigate how primate cortical neurons resolve this tradeoff. While primary basal dendrite number is relatively conserved across mouse cortical areas, it increases systematically along the primate cortical hierarchy. Using biophysically constrained multicompartment modeling, we identify a key dimensionless control parameter—the dendritic-to-somato-apical specific membrane resistance ratio, $\rho \equiv R_{m,\mathrm{dend}}/R_{m,\mathrm{sa}}$, that governs the coupling between dendritic morphology and the effective somatic time constant τ in the weak axial-attenuation limit. In mouse-like neurons ($\rho > 1$), adding basal dendrites progressively increases τ and promotes reliable low-pass integration. Conversely, human-like neurons operate near a balanced-resistance regime ($\rho \approx 1$), where resistive reweighting counterbalances morphology-driven capacitive loading. This allows τ to remain nearly invariant despite expanded dendritic topology, shifting the structure-function tradeoff toward faster, more flexible coding with improved high-frequency tracking. Our results reveal a biophysical scaling principle linking evolutionary dendritic architecture to hierarchy-dependent coding, offering potential design rules for preserving temporal bandwidth in dendrite-inspired neuromorphic architectures.

1 Introduction

Understanding the computational properties of single neurons is a central problem in neuroscience and a prerequisite for explaining biological intelligence. A defining feature of cortical pyramidal neurons is their elaborate dendritic morphology: unlike the point-like units used in most artificial neural networks, they possess spatially structured nonlinear dendritic integration and compartment-specific electrophysiological properties [25, 14, 28]. These structures endow individual neurons with rich computational capabilities. For example, dendrite-inspired multi-branch architectures can improve robustness, temporal processing, and parameter efficiency [46, 11]. Across mammalian evolution, the dendritic architecture becomes progressively more elaborate, with pronounced expansion in primate

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

and human cortex [5]. These observations suggest that biological intelligence may scale not only by increasing the number of neurons, but also by increasing the computational capacity of individual neurons through dendritic expansion.

Yet the temporal consequences of this expansion are paradoxical. At the network level, higher-order cortical areas exhibit progressively slower intrinsic timescales, driven in part by recurrent excitation and excitatory–inhibitory balance gradients [9, 55, 31]. At the single-cell level, larger neurons across species have lower fire rates, however human pyramidal neurons fire action potentials with faster onset dynamics than their smaller rodent counterparts [17]. These observations pull in opposite directions and leave a fundamental question unresolved: does the membrane time constant governing subthreshold integration timescale grow larger in bigger neurons with more elaborate dendrites? More specifically, does dendritic expansion impose tighter biophysical constraints on the membrane time constant, or can time constant remain stable despite expanded morphology?

To address this question, we first perform a systematic cross-species survey of public datasets of cortical pyramidal neurons. While previous studies have mainly characterized particular species or cortical areas, we integrate measurements across mouse and primate cortices and find that the number of primary basal dendrites provides a robust morphological axis for comparing dendritic elaboration across species and hierarchy. This measure captures major integrative branches receiving intracortical input and is less sensitive than higher-order branching statistics to partial reconstructions or dataset-specific tracing differences. Basal dendrite number remains relatively conserved across mouse cortical areas, but increases systematically along the primate cortical hierarchy, consistent with the expansion of dendritic complexity in primate association cortex [33, 5]. Surprisingly, this increase in dendritic complexity is not accompanied by a corresponding increase in the somatic membrane time constant. By integrating published electrophysiological measurements across species and cortical areas, we find that somatic membrane time constants remain within a relatively stable and fast range despite substantial variation in dendritic morphology [5]. This observation suggests a morpho-temporal decoupling mechanism in biological neurons: dendritic morphology can expand across species and cortical hierarchy without imposing a proportional change in the membrane time constant.

To understand this morpho-temporal decoupling, we developed a biophysically constrained multicompartment model and identified a dimensionless electrotonic control parameter, $\rho = \frac{R_{m,\mathrm{dend}}}{R_{m,\mathrm{sa}}}$, where $R_{m,\mathrm{dend}}$ is the basal-dendritic specific membrane resistance and $R_{m,\mathrm{sa}}$ is the effective somato-apical specific membrane resistance. This ratio determines whether added basal dendrites renormalize the effective somatic time constant or leave it approximately invariant. In mouse-like electrotonic organization, $\rho > 1$, added basal dendrites increase τ and favor reliable low-pass integration. In primate/human-like organization, neurons operate near a balanced-resistance regime, $\rho \approx 1$, where compartmental resistance reweighting counterbalances morphology-driven capacitive loading and preserves fast somatic dynamics despite expanded dendritic topology. Thus, ρ acts as a morpho-temporal coupling variable that determines soma membrane time constant across species.

Finally, we show that these distinct electrotonic regimes have functional consequences for temporal coding. Mouse-like neurons support more stable integration, whereas primate/human-like neurons exhibit more flexible temporal coding and broader coding bandwidth. In a supervised spiking network with learnable time constant, the primate-like profile outperforms the tested alternatives, which provides an evidence for a task-dependent architectural prior.

These results suggest that morpho-temporal decoupling is not merely a passive biophysical phenomenon, but a computational mechanism that allows dendritic complexity and temporal coding flexibility to scale together.

Overall, our study reveals a morpho-temporal decoupling principle in biological neurons: (1) identify primary basal dendrite number as a robust structural axis for cross-species and cross-area dendritic scaling; (2) show that morphology–time constant coupling is controlled by a compact electrotonic ratio rather than dendrite number alone; and (3) demonstrate that resistance-matched dendritic expansion supports flexible temporal coding, suggesting a design guideline for temporally efficient dendrite-inspired neural architectures.

2 Related Work

Timescales versus membrane time constants. Intrinsic timescales usually refer to the autocorrelation decay of spiking activity or population responses, forming a cortical hierarchy from fast sensory

areas to slower association and prefrontal areas [43, 8]. These gradients are typically explained by recurrent excitation, excitation–inhibition interactions, inter-areal coupling, synaptic kinetics, and slow network modes [9, 55, 31]. It is noteworthy that the model in these works is rate-based with the same τ for all excitatory cortical units. By contrast, the membrane time constant studied here is a single-cell property that describes the relaxation of subthreshold somatic voltage after small current perturbations and determines synaptic temporal filtering before spike generation [5]. Thus, activity timescales and membrane time constants are related but distinct: the former reflect emergent circuit-level temporal correlations, whereas the latter reflect effective electrotonic filtering at the single-neuron level. Our work focuses on the latter, asking whether dendritic expansion changes the effective somatic membrane time constant.

Dendritic computation. Dendrites have complex morphologies and nonlinear response that allow them to carry substantial single-neuron computations. Branch-specific nonlinear integration, NMDA- and Ca^{2+}-dependent local spikes, sequence selectivity, input multiplexing, and compartmentalized plasticity enable pyramidal neurons to perform computations beyond point-neuron models [27, 29, 7, 30, 53]. These mechanisms have motivated two-layer or deep-network-like models of single-neuron computation, as well as artificial and neuromorphic architectures that use branch-like compartments or dendritic dynamics to improve expressivity, temporal processing, robustness, and parameter efficiency [46, 6, 20, 11, 10, 60]. However, dendritic expansion also increases membrane capacitance and alters axial coupling and electrotonic filtering, thereby potentially changing the effective somatic membrane time constant [48, 17, 23]. Our work asks whether dendritic computational capacity can expand without compromising fast somatic response dynamics.

3 Primate basal dendrites expand along the cortical hierarchy while mouse dendrites do not

Basal dendrite number varies markedly between rodents and primates, yet whether this divergence reflects a conserved principle or a hierarchy-dependent reorganization remains unclear. We first quantify basal-dendrite number across cortical areas of rodents (mouse) and primates (monkey) using publicly available morphological datasets from the Allen Institute for Brain Science [2] and NeuroMorpho [52]. To ensure cross-dataset comparability, we use the number of primary basal dendrites emerging from the soma, a topology metric robust to partial reconstructions and secondary-branching differences.

In mice, layer 2/3 (L2/3) pyramidal neurons exhibit a remarkably conserved basal dendrite count (typically 4–6) across six cortical areas spanning the anatomical hierarchy [24, 50] (V1, $n = 91$[2]; AL, $n = 29$ [2]; PM, $n = 11$[2]; PPC, $n = 19$ [32]; PFC, $n = 26$ [4]; ACC, $n = 29$ [56, 61]; see Supplementary A for full names, Fig. 1A). Basal dendrite number shows no systematic grading across areas ($p = 0.52$, Dunn's test), indicating weak cortical differentiation at this topological level.

In contrast, primate L2/3 exhibits a pronounced hierarchy-dependent organization. Across seven areas (V1, $n = 29$[16, 15]; V2, $n = 23$[16, 15]; V4, $n = 32$[16, 15]; IPS, $n = 59$[42]; PFC, $n = 31$[15]; TS, $n = 43$[13]; ACC, $n = 12$[38]), basal dendrite number increases generally with cortical hierarchical level (Fig. 1B, $p < 0.001$, Dunn's test), quantified by the fraction of labeled neurons derived from laminar projection patterns [37, 36, 35] (Basal dendrite number of PFC may be underestimated due to limited sampling of sub-areas). Higher-order association cortex (ACC) exhibits approximately twice the basal dendrite number observed in primary visual cortex [15, 38].

These morphology-level trends are consistent with independent measurements of dendritic spine organization: mouse spine density is largely invariant across areas, whereas primate association cortex shows dramatic increases in total spine number (up to an order of magnitude), mirroring hierarchical basal-dendrite expansion [3]. Together, these data establish primary basal dendrite number as a structural axis that is conserved within mouse cortex but scales with hierarchical level in primates, raising the question of whether this morphological divergence is accompanied by a corresponding divergence in somatic temporal dynamics.

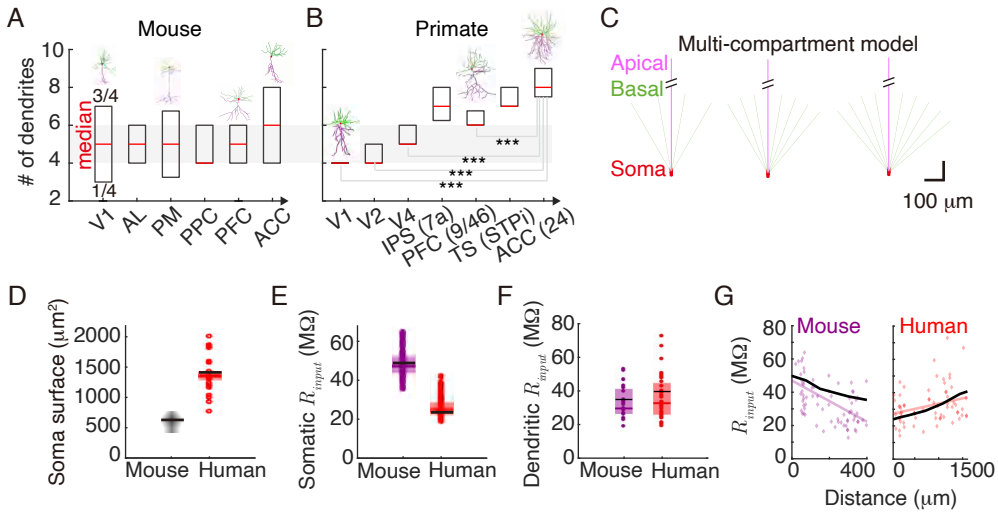


Figure 1: (A-B) Basal dendrite counts across cortical areas in mouse (A) and monkey (B). Boxes indicate the interquartile range, and red bars denote medians. Asterisks mark significant pairwise differences (Dunn's test on mean-rank differences with Dunn–Sidak correction). For clarity, we report comparisons only against the highest-level area ACC. (C) Biophysically constrained multi-compartment models with 4 (left), 7 (middle), and 9 (right) basal dendrites. Apical dendrite length is 900 μm for mouse and 1800 μm for primate. (D–G) Reported morphological and biophysical measurements for mouse (purple) and primate (red), with model values in black: (D) soma surface area; (E) somatic input resistance; (F) dendritic input resistance; (G) distance-dependent apical input resistance from the soma.

4 A unity-resistance principle decouples dendritic expansion from somatic slowing

4.1 Multicompartment model constrained by cross-species electrophysiology

Independent of the morphological divergence documented above, published recordings reveal that mouse and human pyramidal neurons also differ in their passive electrical organization: human neurons exhibit larger soma surface area, lower somatic input resistance, and a distinct distance-dependent profile of dendritic input resistance [5]. Does this resistive reorganization merely accompany dendritic expansion, or does it actively shape how additional branches influence somatic dynamics? To dissect the interplay between morphological scaling and compartmental electrical properties, we construct a biophysically grounded multicompartment model in which basal dendrite number is the primary structural degree of freedom, while somatic and apical electrotonic properties are independently constrained by published mouse and primate/human measurements.

In brief, each neuron comprises a soma, an axon, a single apical dendrite, and a variable number of basal dendrites. Compartments are modeled as uniform cylinders, and higher-order branching is omitted to isolate the effect of primary basal topology. Complete model parameter settings are presented in supplementary B. Basal dendrite counts are set to the median values of the observed distributions (Fig. 1A–B). Beyond basal number, we constrain soma surface area (Fig. 1D), somatic input resistance (Fig. 1E), dendritic input resistance (Fig. 1F), and the distance-dependent apical input-resistance profile (Fig. 1G), and verify that simulated outputs fall within experimentally reported ranges [5]. We adopt physiological measurements from Beaulieu-Laroche et al. in both mouse and primate/human models since external L2/3 constraints remain incomplete [5]. For brevity, we denote primate/human neurons as human-like neurons.

Active conductance is included only to generate spiking under fluctuating drive and to quantify reliability, flexibility and coding metrics. Ionic mechanisms are adopted from experimentally characterized mouse CA1 pyramidal neurons [39] (see Supplementary for channel densities B.2). Importantly, the morphology–τ scaling and resistive balance we report below are governed by passive cable properties rather than the particular choice of voltage-gated conductance.

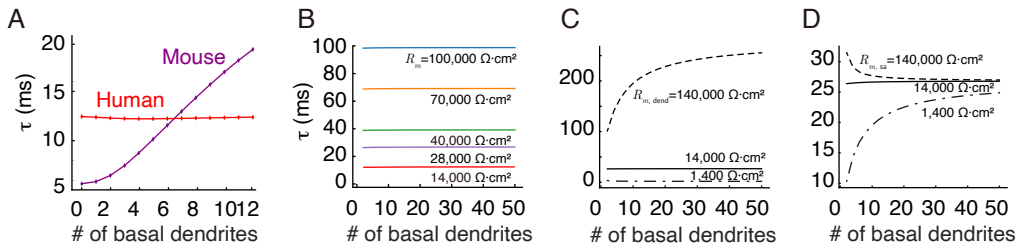


Figure 2: (A) Simulated effective somatic time constant τ versus basal dendrite number for mouse (purple) and human (red) models. (B) τ under homogeneous membrane properties while varying global R_m. (C) As in (B), but varying basal-dendrite R_m. (D) As in (B), but varying somatic R_m. R_m denotes a local membrane property.

4.2 Mouse neurons slow with added dendrites while human-like neurons do not

What biophysical property underlies the morphological divergence identified in Section 3? A natural candidate is the effective somatic membrane time constant τ, which governs the subthreshold timescale over which the soma integrates synaptic inputs. We quantify the τ in both models by fitting a single exponential to the somatic voltage response to a -100 pA hyperpolarizing current step under current-clamp simulation. We observe a pronounced species-dependent relationship between basal dendrite number and τ (Fig. 2A). In human-like model, τ is nearly invariant across a wide range of basal dendrite counts, with a mean of ~ 12.4 ms, consistent with somatic recordings from human L5 pyramidal neurons reporting mean $\tau \approx 12.0$ ms ($n = 39$ [5]). Human L2/3 pyramidal neurons show comparable values (13.7 ms, $n = 56$ [41]; 12.0 ms, $n = 25$ [12]; and ~ 11 ms by calculation [45]). While some studies report larger τ values, potentially reflecting age and inter-individual differences, there is no consistent evidence for strong area-to-area differences in τ under comparable conditions [21].

By contrast, mouse τ varies strongly with basal topology: as basal dendrite number increases over the same range, τ rise from 5.5 ms to 19.9 ms (262% increase). Within the typical mouse L2/3 range (4–6 dendrites), the model yields mean $\tau \sim 11.0$ ms, close to the Allen Institute database mean (~ 12.4 ms; data not shown). Mouse Layer 5 (L5) neurons typically have larger arbors than L2/3 [22] (5–8 basal dendrites from 25–75%); correspondingly, our model increases τ from 11.0 to 14.7 ms, consistent with the trend toward ~ 19.5 ms in Allen Institute L5 measurements. The remaining gap likely reflects the omission of secondary branching in our simplified morphology. These results expose a stark asymmetry: in mouse-like electrotonic organization, τ is tightly coupled to dendritic morphology, whereas in human-like organization it is largely decoupled. We trace this pattern to a single dimensionless control parameter in the following section.

4.3 A closed-form solution reveals ρ as an approximate morpho-temporal switch

We derive a closed-form expression for τ as a function of basal dendrite number N, which describes whether added dendrites slow the soma or leave it unchanged. Each basal dendrite couples to the soma through an axial resistance R_A and presents a membrane admittance $Y_{\text{dend}}(s) = (g_{\text{dend}} + s\,C_{\text{dend}})/(1 + R_A(g_{\text{dend}} + s\,C_{\text{dend}}))$ in the Laplace domain. The total somatic input admittance is then

$$Y_{\text{total}}(s) \;=\; g_{\text{sa}} + s\,C_{\text{sa}} + N \cdot Y_{\text{dend}}(s), \tag{1}$$

where g_{sa} and C_{sa} are the lumped somato-apical conductance and capacitance. Linearizing around $s \to 0$ yields an effective conductance $G_{\text{eff}} = Y_{\text{total}}(0)$ and an effective capacitance $C_{\text{eff}} = \partial Y_{\text{total}}/\partial s|_{s=0}$ (see Supplementary F for full derivation), giving

$$\tau(N) \;=\; \frac{C_{\text{eff}}}{G_{\text{eff}}} \;=\; \frac{C_{\text{sa}} + N \cdot \dfrac{C_{\text{dend}}}{(1 + R_A\, g_{\text{dend}})^2}}{g_{\text{sa}} + \dfrac{N\, g_{\text{dend}}}{1 + R_A\, g_{\text{dend}}}}. \tag{2}$$

First, when $N = 0$, τ reduces to the somato-apical time constant $\tau_{\text{sa}} = C_{\text{sa}}/g_{\text{sa}}$. Second, as $N \to \infty$, τ converges to $\tau_{\text{dend}}/(1 + g_{\text{dend}}\, R_A) \approx \tau_{\text{dend}}$, independent of the soma. Critically, the squared $(1 + R_A\, g_{\text{dend}})^2$ in the capacitance term means that the axial resistance attenuates the capacitive load of each dendrite more strongly than it attenuates its conductive load. This asymmetry is what makes

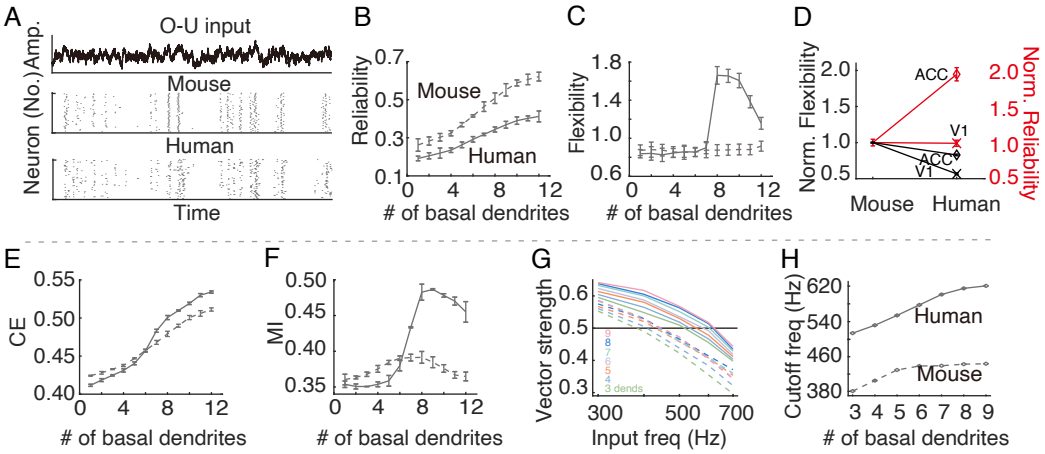


Figure 3: Response properties of mouse and human neuronal models. (A) Example O–U inputs (top), spike rasters from mouse (middle) and human (bottom) neurons with 12 basal dendrites. (B) Reliability quantified by cross-trial covariance. (C) Flexibility quantified by the coefficient of variation of inter-spike intervals (CV_{isi}); error bars indicate s.e.m. (D) Cross-area comparison between mouse and human. (E) Coding efficiency (CE). (F) Mutual information (MI). In (B), (E), and (F), error bars denote s.d. across five dendritic O–U input realizations. (G) Mean vector strength for fast tracking in mouse (dashed) and human (solid) models; cutoff frequency is defined when vector strength drops below 0.5, shown for mouse and human (H).

τ sensitive to the ratio of dendritic and somato-apical resistive properties, rather than to membrane area alone.

Within the reduced model and for fixed per-branch electrical parameters, Eq. (2) shows that the N-dependence of τ is governed jointly by the dendritic-to-somato-apical resistance ratio and axial attenuation. The specific membrane capacitance is commonly approximated as spatially uniform across somatic and dendritic compartments[19, 49]. Under this assumption, $\rho \equiv R_{m,\mathrm{dend}}/R_{m,\mathrm{sa}} \approx \tau_{\mathrm{dend}}/\tau_{\mathrm{sa}}$ in weak axial-attenuation limit. Near this limit, the numerator and denominator of Eq. (2) grow in approximately matched proportions with N, leaving τ weakly dependent on dendrite number (Fig. 2B). More generally, τ increases when $\rho > 1$ and decreases when $\rho < 1$ (Fig. 2C–D). We refer $\rho \approx 1$ as an approximate unity-resistance regime, in which added dendrites contribute capacitive and conductive loads in nearly matched proportions, thereby reducing morphology–τ coupling.

5 Balanced resistance expands temporal bandwidth without sacrificing reliability.

5.1 Distinct electrotonic regimes produce divergent coding profiles

Since ρ controls the morphology–τ coupling, it should also shape the coding repertoire available to each neuron class. We test this prediction by comparing mouse and human-like models across five complementary metrics spanning reliability, flexibility, information transmission, and temporal bandwidth . We compare the morphology-and-conductance-restricted mouse and human-like models across basal-dendrite number in reliability, flexibility, coding efficiency (CE), mutual information (MI), and high-frequency tracking (see Supplementary Sections C, D, and E). Reliability is quantified using dual Ornstein–Uhlenbeck (O–U) current injections: a dendritic drive representing shared feedforward input and a weaker somatic drive representing neuron-specific noise. Within each species, input strength is fixed and scaled so that the two-basal-dendrite model fires at 15 Hz; results are averaged over five O–U realizations, and reliability is measured by trial-to-trial covariance. Flexibility is quantified by the inter-spike-interval coefficient of variation (CV_{isi}) under a single somatic O–U drive normalized to yield 2 Hz (see Supplementary D for other rate condition).

Consistent with dendritic low-pass filtering mechanism [26], reliability increases with dendrite number in both models, but mouse model is more reliable across the full range, most prominently at 12 basal dendrites (Fig. 3A–B). This difference is not explained by morphology alone, since mouse neurons scaled to human size retain typical mouse reliability (see supplementary Fig. S2). In contrast, human-like model in the 8–12 dendrite range exhibits increased flexibility via burst spiking ($CV_{\mathrm{isi}} > 1$; Fig. 3C), overlapping the range reported for primate ACC. Burst-mediated signaling is

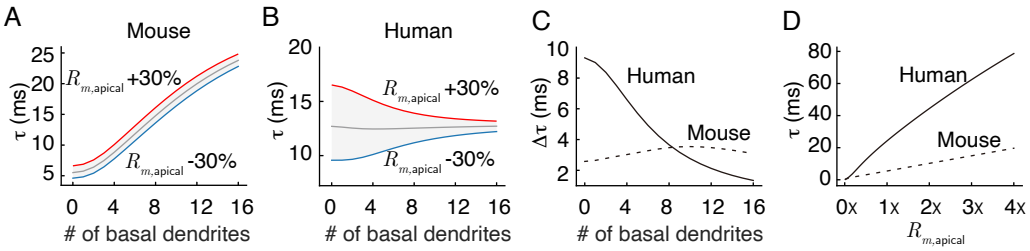


Figure 4: (A) Simulated time constant of mouse neuron with the R_m of apical dendrite ($R_{m,apical}$) perturbed by $\pm 30\%$. (B) Same as A except human neuron. (C) $\Delta\tau$ between the $\pm 30\%$ perturbation of human (solid) and mouse (dash) neuron. (D) Simulated τ with different fold of apical R_m.

implicated in information transfer in cingulate and prefrontal circuits [57], suggesting a biophysical route to enhanced coding flexibility in higher-order areas. While mouse neurons can also burst in vivo, it is often channel-driven rather than morphology-driven [59]; consistently, size-expanded mouse model requires $\gtrsim 12$ basal dendrites to substantially increase flexibility (see supplementary Fig. S3). Across homologous regions, human V1 shows flexibility comparable to mouse V1 but only 53% of its reliability; in ACC, reliability is closer (human ACC at 76% of mouse ACC), whereas flexibility is $1.82\times$ higher (Fig. 3D).

Across dendrite number, mouse models with ≤ 6 basal dendrites exhibit higher CE (Fig. 3E) and MI (Fig. 3F), whereas human model gains a clear advantage beyond six dendrites, corresponding to higher-order cortices. MI gains largely reflect burst-associated rate increases and an enlarged spike pool for encoding, whereas CE is largely rate independent, indicating a mechanism beyond firing-rate changes. These trends are robust to varying $\tau_{\mathrm{O-U}}$ from 3 to 10 ms (see supplementary Fig. S4). We quantify high-frequency tracking using normalized vector strength [17], computed from the first spike within an 8 ms window to exclude burst clustering. The cutoff frequency is denoted as the frequency where vector strength falls below 0.5. Human-like model presents a broader temporal bandwidth [54, 17] (Fig. 3G), with cutoff frequency increasing with dendrite number (Fig. 3H). In mouse model, cutoff frequency saturates for ≥ 6 dendrites, indicating diminishing returns from additional branches specifically for fast temporal tracking, and provides a mechanistic basis for the stabilized basal-dendrite number near six in mouse cortex. In summary, the electrotonic regime sets a structure–function tradeoff: mouse-like neurons ($\rho > 1$) convert added dendrites primarily into slower, more reliable integration, whereas human-like neurons ($\rho \approx 1$) convert them into broader temporal bandwidth, higher coding capacity, and greater flexibility without paying a temporal cost.

6 Hierarchy-dependent τ heterogeneity: from biophysics to network function

6.1 Inverse gradient of τ variability

The dimensionless resistance ratio ρ is the key control parameter linking basal dendrite number to the effective somatic time constant τ; thus, active modulation of ρ provides a direct mechanism to tune τ over a biologically relevant range. In artificial neural networks, adaptive control of τ can enhance computational flexibility [18], motivating the possibility that regulated τ heterogeneity may likewise support biological computation.

In cortex, ρ can be modulated via top-down regulation of apical conductance through two established mechanisms: (i) concurrent E/I conductance changes induced by top-down inputs [40]; and (ii) distance-dependent HCN expression along apical dendrites, where suppressing I_h increases R_m and prolongs τ [34]. Guided by these observations, we simulated $\pm 30\%$ modulation of apical membrane resistance $R_{m,\mathrm{apical}}$.

In mouse-like model, varying $R_{m,\mathrm{apical}}$ preserves the positive dependence of τ on dendritic complexity (Fig. 4A), producing mainly a baseline shift ($\Delta\tau = 2.2$ ms; slope $\kappa = 0.03$ ms/dendrite). In human-like model, the effect is qualitatively different: sparse basal arbors exhibit large τ variability (consistent with primary areas), whereas extensive basal arbors maintain a stable τ (consistent with higher-order areas; Fig. 4B). Quantitatively, $\Delta\tau$ declines steeply as basal dendrite number increases from 0 to 16 (9.2 ms to 1.8 ms; -0.46 ms/dendrite; Fig. 4C), predicting larger τ heterogeneity in primate primary sensory cortex and overall reduction in heterogeneity in higher-order association cortex.

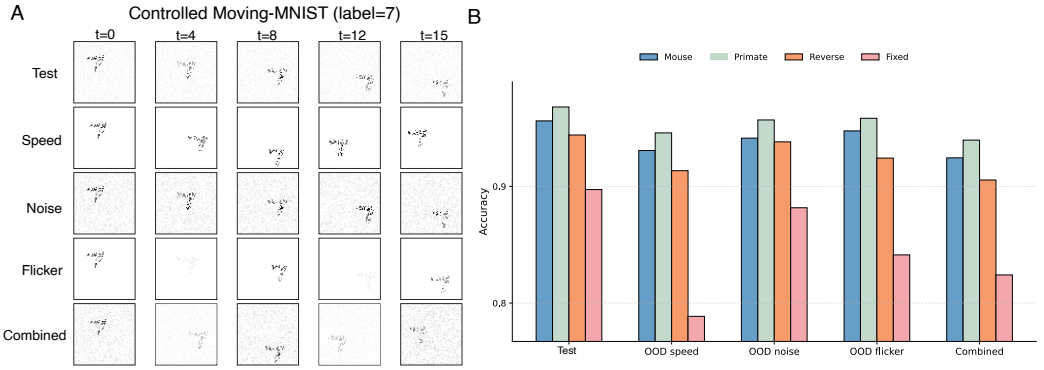


Figure 5: **Robust temporal representation in Moving-MNIST.** (A) Example sequences from Controlled Moving-MNIST showing the standard test condition and OOD perturbations in speed, noise, flicker, and their combination. (B) Accuracy of four network variants across these conditions. Among the four tested variance profiles, the primate-like profile achieves the highest accuracy under the evaluated in-distribution and out-of-distribution conditions.

These predictions are supported by macaque recordings: V1 shows larger inter-neuron τ variability (22.6$\pm$2.4 ms [21]; 19.1$\pm$3.6 ms [1]) than lateral prefrontal cortex (LPFC: 18.1$\pm$1.1 ms [1]). In mice, by contrast, τ distributions have similar variance across regions (V1: 27.6$\pm$1.7 ms; frontal cortex: 36.7$\pm$2.5 ms [21]; reported differences are not significant). Mechanistically, this divergence follows from higher effective $R_{m,\mathrm{apical}}$ in primate/human-like neurons (Fig. 4D). Although apical input resistance appears similar across species (Fig. 1E), it is a whole-cell measure: in rodents, higher somatic $R_{m,\mathrm{soma}}$ can partially mask low apical $R_{m,\mathrm{apical}}$. To match these constraints, the primate/human-like model uses $R_{m,\mathrm{apical}}$ that increases with distance from the soma (distal tip: 70 kΩcm^2), whereas the mouse-like model uses a steeply decreasing profile (distal tip: 1.4 kΩcm^2), consistent with HCN distributions [23, 5]. Consequently, primate-like architectures exhibit greater sensitivity of τ to resistance modulation, enabling hierarchy-dependent stabilization of τ with increasing basal complexity.

6.2 Hierarchical τ heterogeneity supports robust temporal representation

The mechanism identified above suggests that dendritic morphology can be decoupled from the mean effective time constant, while time-constant variability remains hierarchically reorganized. Mouse-like neurons show comparable variability across cortical areas, whereas primate/human-like neurons exhibit a progressive contraction of variability along the cortical hierarchy. This observation raises a functional question: why would the cortex preserve a similar mean temporal integration scale while redistributing time-constant variability across hierarchical levels?

We hypothesize that this redistribution of time-constant variability supports hierarchical temporal processing. Heterogeneity in early sensory areas may help encode stimuli varying over multiple temporal scales, such as motion, flicker, and noise-corrupted input, whereas reduced variability in higher cortical areas may stabilize abstract representations by limiting unnecessary temporal dispersion. Thus, the hierarchical contraction of time-constant variability may balance temporal flexibility in early processing with representational stability in deeper stages.

To test this idea, we train a compact four-layer convolutional spiking network on a temporally structured visual classification task (details in Supplementary G). The network uses PLIF neurons [18], which allow the membrane time constant to be parameterized and learned within each layer. Across model variants, we keep the mean time constant matched but impose different profiles of time-constant heterogeneity. The *primate* profile assigns larger time-constant variance to early layers and progressively smaller variance to deeper layers, mirroring the primate/human-like pattern. The *reverse* profile uses the opposite arrangement as a control. The *mouse* profile uses matched variance across layers, and the *fixed* profile removes meaningful time-constant heterogeneity by collapsing all channels toward a common value. The total variance budget is matched across the primate, reverse, and mouse profiles, ensuring that the comparison isolates the effect of hierarchical variance allocation rather than the overall amount of heterogeneity.

The task is designed as a Moving MNIST video-classification problem with temporal perturbations spanning multiple timescales. Each sample consists of a $T = 16$ frame video in which one MNIST digit moves linearly on a 64×64 canvas and reflects at the boundaries. The label is the digit identity, as in standard MNIST, but the input contains task-irrelevant temporal factors, including direction, speed, frame-wise luminance flicker, and additive Gaussian noise. To increase the demand for temporal integration, foreground pixels within the digit bounding box are randomly retained with low probability, producing sparse and fragmented visual evidence in each frame (Fig. 5A). The network is therefore required to classify digit identity from a degraded and dynamically changing visual stream rather than from a clean static image. We further evaluate out-of-distribution generalization using held-out speeds, flicker frequencies, noise levels, and combined temporal perturbations that are absent during training.

Within this controlled four-layer feedforward PLIF network and the Moving-MNIST perturbations considered here, the primate-like profile achieves the strongest out-of-distribution generalization among the four tested variance profiles, followed by the mouse-like profile, whereas the reverse and fixed profiles perform less well. These results support hierarchy-dependent time-constant variance allocation as a useful architectural prior for temporal robustness in this setting. Its benefit may depend on the temporal statistics of the task, network depth, recurrent connectivity, learning objective, and the layer-wise distribution of temporal features.

7 Summary and Conclusions

We identify a morpho-temporal decoupling principle in primate/human cortical pyramidal neurons: large increases in dendritic morphological degrees of freedom can coexist with a fast, nearly invariant somatic time constant τ. Combining cross-species constraints with biophysically grounded multicompartment modeling, we show that compartment-specific redistribution of specific membrane resistance can organize neurons into an approximate unity-resistance condition, captured by the dimensionless ratio $\rho \equiv R_{m,\mathrm{dend}}/R_{m,\mathrm{sa}}$ with $\rho \approx 1$. In this regime, morphology-driven capacitive loading is counterbalanced by resistive reweighting across compartments, rendering τ weakly sensitive to basal-dendrite expansion.

From a physics perspective, ρ is a compact dimensionless control parameter for the transition between morphology–timescale coupling and decoupling: it quantifies the competition between capacitive loading from added dendritic degrees of freedom and electrotonic reweighting via membrane-resistance redistribution, yielding an effective renormalization of the somatic time constant. Active dendritic nonlinearities (e.g., NMDA and Ca^{2+} spikes) can further modulate effective resistance and integration time scales; our results therefore provide a passive baseline and predict that active effects will matter most insofar as they shift the effective resistive partitioning away from or toward $\rho \approx 1$.

The same principle suggests a direct link to dendrite-inspired artificial neural networks and neuromorphic design: in multi-branch units, ρ acts as a tunable ratio that controls whether added branches mainly increase effective integration (stability/reliability) or preserve fast time constants while expanding representational degrees of freedom. The emergence of a unity-resistance regime in primate/human-like neurons further invites an evolutionary interpretation: selection for maintaining temporal bandwidth under increasing dendritic complexity could favor organization near $\rho \approx 1$, potentially aided by developmental co-regulation of compartmental membrane properties. These ideas yield testable predictions, including hierarchy-dependent patterns of τ heterogeneity in primate cortex (Fig. 4) and compartment-specific resistance differences that track dendritic complexity.

Functionally, decoupling allows additional morphological degrees of freedom to be expressed as flexible response modes while preserving fast dynamics, improving coding efficiency and expanding high-frequency tracking bandwidth. By contrast, mouse-like electrotonic organization exhibits stronger morphology–τ coupling; together with the narrow distribution of mouse basal dendrite number, this favors reliable integration but can limit high-frequency tracking as dendritic load increases. Overall, ρ provides a minimal biophysical handle on whether dendritic expansion primarily increases integration (reliability) or preserves coding flexibility and high-frequency bandwidth via morpho-temporal decoupling.

Limitations Our multicompartment models use simplified cylindrical morphologies with uniform basal dendrites, omitting higher-order branching, spine density variation, and dendritic tapering. While this abstraction isolates the role of primary basal topology and ρ, it underestimates the electrotonic heterogeneity present in fully reconstructed neurons, and future work should validate the

unity-resistance prediction in detailed morphological reconstructions. The theoretical framework is built on passive cable properties, limiting the theory application to small-scale subthreshold activity. The active dendritic conductances such as NMDA spikes, Ca^{2+} plateaus, and HCN channels will shift the effective resistive partitioning and could move neurons toward or away from $\rho \approx 1$. Finally, our spiking network uses a simplified feedforward architecture; extending this analysis to recurrent circuits with synaptic plasticity and neuromodulation will be important for linking our findings to broader studies such as neural timescales. Accordingly, the network results should be interpreted as evidence for a useful inductive bias in the tested setting, rather than as demonstrating an optimal variance profile for recurrent networks, other tasks, or other learning objectives.

Funding Funding in direct support of this work: the Science and Technology Innovation 2030 - Brain Science and Brain-Inspired Intelligence Project (2021ZD0201301), the National Natural Science Foundation of China (92570202, 12201125), Shanghai Municipal Science and Technology (24JS2810400 and 21XD1400400) and Shanghai Institute for Mathematics and Interdisciplinary Sciences (SIMISID-2025-NC).

Acknowledgments The computations in this research were performed using the CFFF platform of Fudan University.

Competing interests The authors declare no competing financial interests.

References

[1] J. M. Amatrudo, C. M. Weaver, J. L. Crimins, P. R. Hof, D. L. Rosene, and J. I. Luebke. Influence of Highly Distinctive Structural Properties on the Excitability of Pyramidal Neurons in Monkey Visual and Prefrontal Cortices. *Journal of Neuroscience*, 32(40):13644–13660, Oct. 2012. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2581-12.2012.

[2] A. M. B. Atlas. Available from mouse. brain-map. org. *Allen Institute for Brain Science*, 2011.

[3] I. Ballesteros-Yáñez, R. Benavides-Piccione, GN. Elston, R. Yuste, and J. DeFelipe. Density and morphology of dendritic spines in mouse neocortex. *Neuroscience*, 138(2):403–409, 2006.

[4] I. Ballesteros-Yáñez, R. Benavides-Piccione, J.-P. Bourgeois, J.-P. Changeux, and J. DeFelipe. Alterations of cortical pyramidal neurons in mice lacking high-affinity nicotinic receptors. *Proceedings of the National Academy of Sciences*, 107(25):11567–11572, 2010.

[5] L. Beaulieu-Laroche, N. J. Brown, M. Hansen, E. H. Toloza, J. Sharma, Z. M. Williams, M. P. Frosch, G. R. Cosgrove, S. S. Cash, and M. T. Harnett. Allometric rules for mammalian cortical layer 5 neuron biophysics. *Nature*, 600(7888):274–278, 2021.

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

[7] T. Branco, B. A. Clark, and M. Häusser. Dendritic Discrimination of Temporal Input Sequences in Cortical Neurons. *Science*, 329(5999):1671–1675, Sept. 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1189664. URL `https://www.science.org/doi/10.1126/science.1189664`.

[8] S. E. Cavanagh, L. T. Hunt, and S. W. Kennerley. A Diversity of Intrinsic Timescales Underlie Neural Computations. *Frontiers in Neural Circuits*, 14:615626, Dec. 2020. ISSN 1662-5110. doi: 10.3389/fncir.2020.615626. URL `https://www.frontiersin.org/articles/10.3389/fncir.2020.615626/full`.

[9] R. Chaudhuri, K. Knoblauch, M.-A. Gariel, H. Kennedy, and X.-J. Wang. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron*, 88(2):419–431, 2015.

[10] S. Chavlis and P. Poirazi. Dendrites endow artificial neural networks with accurate, robust and parameter-efficient learning. *Nature Communications*, 16, Jan. 2025. doi: 10.1038/s41467-025-56297-9.

[11] S. D'Agostino, F. Moro, T. Torchet, Y. Demirağ, L. Grenouillet, N. Castellani, G. Indiveri, E. Vianello, and M. Payvand. DenRAM: Neuromorphic dendritic architecture with RRAM for efficient temporal processing with delays. *Nature communications*, 15(1):3446, Apr. 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-47764-w.

[12] Y. Deitcher, G. Eyal, L. Kanari, M. B. Verhoog, G. A. Atenekeng Kahou, H. D. Mansvelder, C. P. J. de Kock, and I. Segev. Comprehensive Morpho-Electrotonic Analysis Shows 2 Distinct Classes of L2 and L3 Pyramidal Neurons in Human Temporal Cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 27(11):5398–5414, Nov. 2017. ISSN 1460-2199 1047-3211. doi: 10.1093/cercor/bhx226.

[13] H. Duan, SL. Wearne, JH. Morrison, and PR. Hof. Quantitative analysis of the dendritic morphology of corticocortical projection neurons in the macaque monkey association cortex. *Neuroscience*, 114(2):349–359, 2002.

[14] G. N. Elston, R. Benavides-Piccione, and J. DeFelipe. The pyramidal cell in cognition: A comparative study in human and monkey. *The Journal of Neuroscience*, 21(17):RC163, 2001.

[15] G. N. Elston, T. Oga, and I. Fujita. Spinogenesis and pruning scales across functional hierarchies. *Journal of Neuroscience*, 29(10):3271–3275, 2009.

[16] G. N. Elston, T. Oga, T. Okamoto, and I. Fujita. Spinogenesis and pruning from early visual onset to adulthood: an intracellular injection study of layer iii pyramidal cells in the ventral visual cortical pathway of the macaque monkey. *Cerebral Cortex*, 20(6):1398–1408, 2010.

[17] G. Eyal, H. D. Mansvelder, C. P. de Kock, and I. Segev. Dendrites impact the encoding capabilities of the axon. *Journal of Neuroscience*, 34(24):8063–8071, 2014.

[18] W. Fang, Z. Yu, Y. Chen, T. Masquelier, T. Huang, and Y. Tian. Incorporating Learnable Membrane Time Constant to Enhance Learning of Spiking Neural Networks. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2641–2651, 2021. ISSN 15505499. doi: 10.1109/ICCV48922.2021.00266.

[19] L. J. Gentet, G. J. Stuart, and J. D. Clements. Direct measurement of specific membrane capacitance in neurons. *Biophysical journal*, 79(1):314–320, 2000.

[20] A. Gidon, T. A. Zolnik, P. Fidzinski, F. Bolduan, A. Papoutsi, P. Poirazi, M. Holtkamp, I. Vida, and M. E. Larkum. Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science (New York, N.Y.)*, 367(6473):83–87, Jan. 2020. ISSN 1095-9203 0036-8075. doi: 10.1126/science.aax6239.

[21] J. P. Gilman, M. Medalla, and J. I. Luebke. Area-specific features of pyramidal neurons—a comparative study in mouse and rhesus monkey. *Cerebral Cortex*, 27(3):2078–2094, 2017.

[22] N. W. Gouwens, S. A. Sorensen, J. Berg, C. Lee, T. Jarsky, J. Ting, S. M. Sunkin, D. Feng, C. A. Anastassiou, E. Barkan, et al. Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature neuroscience*, 22(7):1182–1195, 2019.

[23] M. T. Harnett, J. C. Magee, and S. R. Williams. Distribution and function of HCN channels in the apical dendritic tuft of neocortical pyramidal neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(3):1024–1037, Jan. 2015. ISSN 1529-2401 0270-6474. doi: 10.1523/JNEUROSCI.2813-14.2015.

[24] J. A. Harris, S. Mihalas, K. E. Hirokawa, J. D. Whitesell, H. Choi, A. Bernard, P. Bohn, S. Caldejon, L. Casal, A. Cho, et al. Hierarchical organization of cortical and thalamic connectivity. *Nature*, 575(7781):195–202, 2019.

[25] L. He, Y. Xu, W. He, Y. Lin, Y. Tian, Y. Wu, W. Wang, Z. Zhang, J. Han, Y. Tian, B. Xu, and G. Li. Network model with internal complexity bridges artificial intelligence and neuroscience. *Nature Computational Science*, 4(8):584–599, Aug. 2024. ISSN 2662-8457. doi: 10.1038/s43588-024-00674-9. URL `https://www.nature.com/articles/s43588-024-00674-9`.

[26] B. Hutcheon and Y. Yarom. Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends in Neurosciences*, 23(5):216–222, 2000. ISSN 0166-2236. doi: 10.1016/S0166-2236(00)01547-2.

[27] M. Häusser and B. Mel. Dendrites: bug or feature? *Current Opinion in Neurobiology*, 13 (3):372–383, June 2003. ISSN 09594388. doi: 10.1016/S0959-4388(03)00075-8. URL `https://linkinghub.elsevier.com/retrieve/pii/S0959438803000758`.

[28] E. M. Kasper, A. U. Larkman, J. Lübke, and C. Blakemore. Pyramidal neurons in layer 5 of the rat visual cortex. I. Correlation among cell morphology, intrinsic electrophysiological properties, and axon targets. *Journal of Comparative Neurology*, 339(4):459–474, 1994.

[29] M. Larkum. A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex. *Trends in neurosciences*, 36(3):141–151, Mar. 2013. ISSN 1878-108X 0166-2236. doi: 10.1016/j.tins.2012.11.006.

[30] R. Legenstein and W. Maass. Branch-Specific Plasticity Enables Self-Organization of Nonlinear Computation in Single Neurons. *Journal of Neuroscience*, 31(30):10787–10802, July 2011. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.5684-10.2011. URL `https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.5684-10.2011`.

[31] S. Li and X.-J. Wang. Hierarchical timescales in the neocortex: Mathematical mechanism and biological insights. *Proceedings of the National Academy of Sciences*, 119(6):e2110274119, 2022.

[32] S. Loomba, J. Straehle, V. Gangadharan, N. Heike, A. Khalifa, A. Motta, N. Ju, M. Sievers, J. Gempt, H. S. Meyer, et al. Connectomic comparison of mouse and human cortex. *Science*, 377(6602):eabo0924, 2022.

[33] J. I. Luebke. Pyramidal neurons are not generalizable building blocks of cortical networks. *Frontiers in neuroanatomy*, 11:11, 2017.

[34] J. C. Magee. Dendritic hyperpolarization-activated currents modify the integrative properties of hippocampal CA1 pyramidal neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18(19):7613–7624, Oct. 1998. ISSN 0270-6474 1529-2401. doi: 10.1523/JNEUROSCI.18-19-07613.1998.

[35] N. T. Markov, M. Ercsey-Ravasz, D. C. Van Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy. Cortical high-density counterstream architectures. *Science*, 342(6158):1238406, 2013.

[36] N. T. Markov, M. M. Ercsey-Ravasz, AR. Ribeiro Gomes, C. Lamy, L. Magrou, J. Vezoli, P. Misery, A. Falchier, R. Quilodran, M.-A. Gariel, et al. A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral cortex*, 24(1):17–36, 2014.

[37] N. T. Markov, J. Vezoli, P. Chameau, A. Falchier, R. Quilodran, C. Huissoud, C. Lamy, P. Misery, P. Giroud, S. Ullman, et al. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of comparative neurology*, 522(1):225–259, 2014.

[38] M. Medalla, W. Chang, S. Ibañez, T. Guillamon-Vivancos, M. Nittmann, A. Kapitonava, S. E. Busch, T. L. Moore, D. L. Rosene, and J. I. Luebke. Layer-specific pyramidal neuron properties underlie diverse anterior cingulate cortical motor and limbic networks. *Cerebral Cortex*, 32(10): 2170–2196, 2022.

[39] M. Migliore, D. A. Hoffman, J. C. Magee, and D. Johnston. Role of an A-type K+ conductance in the back-propagation of action potentials in the dendrites of hippocampal pyramidal neurons. *Journal of computational neuroscience*, 7:5–15, 1999.

[40] S. Moldakarimov, M. Bazhenov, and T. J. Sejnowski. Top-down inputs enhance orientation selectivity in neurons of the primary visual cortex during perceptual learning. *PLoS computational biology*, 10(8):e1003770, Aug. 2014. ISSN 1553-7358 1553-734X. doi: 10.1371/journal.pcbi.1003770.

[41] H. Moradi Chameh, S. Rich, L. Wang, F.-D. Chen, L. Zhang, P. L. Carlen, S. J. Tripathy, and T. A. Valiante. Diversity amongst human cortical pyramidal neurons revealed via their sag currents and frequency preferences. *Nature Communications*, 12(1):2497, May 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-22741-9.

[42] S. E. Motley, Y. S. Grossman, W. G. Janssen, M. G. Baxter, P. R. Rapp, D. Dumitriu, and J. H. Morrison. Selective loss of thin spines in area 7a of the primate intraparietal sulcus predicts age-related working memory impairment. *Journal of Neuroscience*, 38(49):10467–10478, 2018.

[43] J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, et al. A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience*, 17(12):1661–1663, 2014.

[44] T. Nevian, M. Larkum, A. Poleg-Polsky, and J. Schiller. Properties of basal dendrites of layer 5 pyramidal neurons: A direct patch-clamp recording study. *Nature neuroscience*, 10:206–14, Mar. 2007. doi: 10.1038/nn1826.

[45] G. Oláh, R. Lákovics, S. Shapira, Y. Leibner, A. Szücs, É. A. Csajbók, P. Barzó, G. Molnár, I. Segev, and G. Tamás. Accelerated signal propagation speed in human neocortical dendrites. *eLife*, 13, Apr. 2025. ISSN 2050-084X. doi: 10.7554/eLife.93781.

[46] P. Poirazi, T. Brannon, and B. W. Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37(6):989–999, Mar. 2003. ISSN 0896-6273. doi: 10.1016/s0896-6273(03)00149-1.

[47] R. Pryluk, Y. Kfir, H. Gelbard-Sagiv, I. Fried, and R. Paz. A tradeoff in the neural code across regions and species. *Cell*, 176(3):597–609, 2019.

[48] W. Rall. Time constants and electrotonic length of membrane cylinders and neurons. *Biophysical journal*, 9(12):1483–1508, 1969.

[49] A. Roth and M. Häusser. Compartmental models of rat cerebellar purkinje cells based on simultaneous somatic and dendritic patch-clamp recordings. *The Journal of physiology*, 535(2): 445–472, 2001.

[50] J. H. Siegle, X. Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T. K. Ramirez, H. Choi, J. A. Luviano, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92, 2021.

[51] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek. Entropy and information in neural spike trains. *Physical Review Letters*, 80(1):197–200, 1998. doi: 10.1103/PhysRevLett.80.197. URL `https://doi.org/10.1103/PhysRevLett.80.197`.

[52] C. Tecuatl, B. Ljungquist, and G. A. Ascoli. Accelerating the continuous community sharing of digital neuromorphology data. *FASEB BioAdvances*, 6(7):207–221, 2024.

[53] B. B. Ujfalussy, J. K. Makara, M. Lengyel, and T. Branco. Global and Multiplexed Dendritic Computations under In Vivo-like Conditions. *Neuron*, 100(3):579–592.e5, Nov. 2018. ISSN 08966273. doi: 10.1016/j.neuron.2018.08.032. URL `https://linkinghub.elsevier.com/retrieve/pii/S0896627318307372`.

[54] C. Verbist, M. G. Müller, H. D. Mansvelder, R. Legenstein, and M. Giugliano. The location of the axon initial segment affects the bandwidth of spike initiation dynamics. *PLOS Computational Biology*, 16(7):e1008087, 2020.

[55] X.-J. Wang. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nature Reviews Neuroscience*, 21(3):169–178, Mar. 2020. ISSN 1471-003X, 1471-0048. doi: 10.1038/s41583-020-0262-x. URL `https://www.nature.com/articles/s41583-020-0262-x`.

[56] J. Winnubst, E. Bas, T. A. Ferreira, Z. Wu, M. N. Economo, P. Edson, B. J. Arthur, C. Bruns, K. Rokicki, D. Schauder, et al. Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. *Cell*, 179(1):268–281, 2019.

[57] T. Womelsdorf, S. Ardid, S. Everling, and T. A. Valiante. Burst firing synchronizes prefrontal and anterior cingulate cortex during attentional control. *Current biology : CB*, 24(22):2613–2621, Nov. 2014. ISSN 1879-0445 0960-9822. doi: 10.1016/j.cub.2014.09.046.

[58] Y. Yu, Y. Shu, and D. A. McCormick. Cortical action potential backpropagation explains spike threshold variability and rapid-onset kinetics. *Journal of Neuroscience*, 28(29):7260–7272, 2008.

[59] F. Zeldenrust, W. J. Wadman, and B. Englitz. Neural coding with bursts—current state and future perspectives. *Frontiers in computational neuroscience*, 12:48, 2018.

[60] H. Zheng, Z. Zheng, R. Hu, B. Xiao, Y. Wu, F. Yu, X. Liu, G. Li, and L. Deng. Temporal dendritic heterogeneity incorporated with spiking neural networks for learning multi-timescale dynamics. *Nature Communications*, 15(1):277, Jan. 2024. ISSN 2041-1723. doi: 10.1038/s41467-023-44614-z. URL `https://www.nature.com/articles/s41467-023-44614-z`.

[61] W. Zhou, A. Kavelaars, and C. J. Heijnen. Metformin prevents cisplatin-induced cognitive impairment and brain damage in mice. *PloS one*, 11(3):e0151890, 2016.

A Basal dendrite datasets collection

Monkey dendritic data is collected from NeuroMorpho (NeuroMorpho.org). The abbreviation of monkey cortical area is: visual cortex (V1, V2, V4), intraparietal sulcus (IPS), prefrontal cortex (PFC), temporal sulcus (TS), anterior cingulate cortex (ACC). Only neurons in layer 2/3 is included in Fig. 1, except the available datasets of TS do not specify the laminar layer.

Mouse dendritic data is collected from Allen Institute for Brain Science [22] and NeuroMorpho [52]. The abbreviation of cortical area in mouse is: primary visual cortex (V1), anterolateral visual area (AL), postero-medial visual area (PM). posterior association area (PPC), prefrontal cortex (PFC), agranular cingulate cortex (ACC). Only neurons in layer 2/3 is included in Fig. 1. We also present the V1, PFC and ACC in Layer 5, as well as CA1 in Fig. S1.

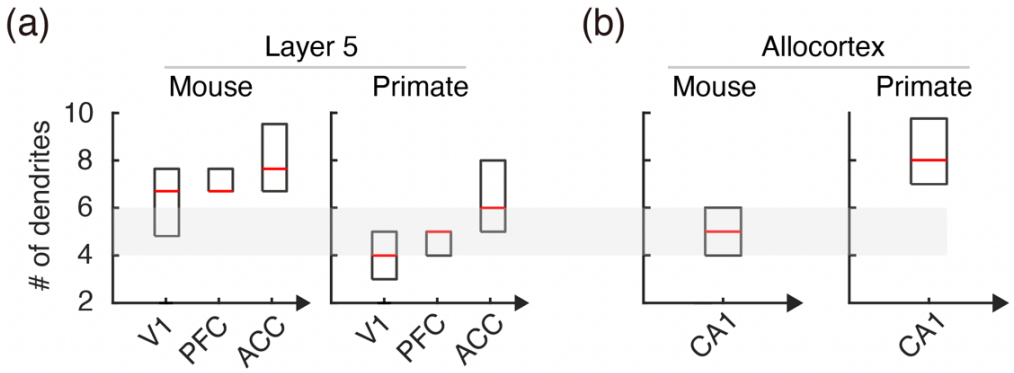


Figure S1: Basal dendrite number in three cortical areas in Layer 5 (a) and Hippocampal CA1 (b).

B Model details

The simulation is performed by NEURON (version 8.2.1) in Python (version 3.9.12). We build the neuronal model based on mouse CA1 multi-compartment model in previous works [39]. The neuron is composed of multiple basal dendrites, one apical dendrite, one soma and one axon. All compartments are modeled as isometric cylinders, and the diameter varies across compartments. Simulation temperature is 37 degrees Celsius, and time step is 0.03 ms.

B.1 Morphology modeling

Soma: The soma of mouse and human neurons accords with the empirical data of layer 5 pyramidal neurons [5]. The soma of mouse neuron is a cylinder with a length of 20 μm and a diameter of 10 μm, yielding somatic surface of nearly 628 μm^2.The soma of human neuron is a cylinder with length of 30 μm and a diameter of 15 μm, yielding somatic surface of nearly 1414 μm^2. Apical dendrite: The apical dendrite is modeled as an isometric cylinder with 5 segments. Beaulieu-Laroche, L. et al. reported the distal apical dendritic recording at the site of 280–400 μm from soma of mouse neurons, and 900–1638 μm from soma of human neurons [5]. Accordingly, we set our mouse apical dendrite with a length of 900 μm and a diameter of 6 μm, and human apical dendrite with a length of 1800 μm and a diameter of 9 μm. The apical diameter is scaled with soma size. Basal dendrite: each basal dendrite has same length (mouse: 500 μm, human: 750 μm). The basal dendrite diameter is scaled by total basal dendrite number to compensate for potential variation for secondary branching (mouse: $0.25 \times N$ μm, human: $0.375 \times N$ μm, with N being the number of basal dendrites, the scale is empirically-derived following the increment of bifurcation and branch number). Axon: The axon of mouse neuron is a cylinder with a length of 50 μm and a diameter of 1 μm, and the soma of human neuron is a cylinder with a length of 75 μm and a diameter of 1.5 μm.

B.2 Biophysical modeling

The membrane resistance (R_m) is restrained to replicate the empirical data [5]. The R_m of soma, axon and basal dendrites of mouse neurons is 40,000 $\Omega \cdot \mathrm{cm}^2$, and 14,000 $\Omega \cdot \mathrm{cm}^2$ for that of human

neurons. To reproduce the distance-dependent profile of apical R_{input} from soma, the R_m of mouse apical dendrite is tuned as 35,000 $\Omega \cdot \text{cm}^2$ at soma tip and 1,400 $\Omega \cdot \text{cm}^2$ at distant tip (900 μm from soma). The decrement factor follows a cubic function, given by $1/(1 + Ax^3)$, where x is normalized distance from soma, 0 at soma and 1 at distant tip. A is scaled factor with value of 24. The R_m of monkey apical dendrite is tuned as 10,000 $\Omega \cdot \text{cm}^2$ at soma tip and 70,000 $\Omega \cdot \text{cm}^2$ at distant tip (1800 μm from soma). The increment factor following a linear function $1/(1 - Ax)$, where A is a scaled factor with value of 0.8.

The configuration of ion channels originate from mouse CA1 neuron [39], and biophysical parameters in the model is primarily referenced from pervious works [17, 44, 58]. In both mouse and human model, the transient Na current density is moderately high in the axon (8,000 $\text{pS}/\mu\text{m}^2$), but low in the soma (400 $\text{pS}/\mu\text{m}^2$) and dendrites (apical: 100 $\text{pS}/\mu\text{m}^2$, basal: 20 $\text{pS}/\mu\text{m}^2$).

Concerning mouse neurons, the delayed rectifier K conductance (g_{Kdr}), has density of 3,000 $\text{pS}/\mu\text{m}^2$ at axon, 100 $\text{pS}/\mu\text{m}^2$ at soma, basal dendrites and apical dendrites. The fast, voltage-activated K conductance (g_{Kv}), has density of 3000 $\text{pS}/\mu\text{m}^2$ at axon, 100 $\text{pS}/\mu\text{m}^2$ at soma and 1 $\text{pS}/\mu\text{m}^2$ at basal dendrites, and a linear gradient from 1 to 51 $\text{pS}/\mu\text{m}^2$ at apical dendrites.

Concerning human neurons, the delayed rectifier K conductance (g_{Kdr}), has density of 4,000 $\text{pS}/\mu\text{m}^2$ at axon, 100 $\text{pS}/\mu\text{m}^2$ at soma, basal dendrites and apical dendrites. The fast, voltage-activated K conductance (g_{Kv}), has density of 4,000 $\text{pS}/\mu\text{m}^2$ at axon, 100 $\text{pS}/\mu\text{m}^2$ at soma and 1 $\text{pS}/\mu\text{m}^2$ at basal dendrites, and a linear gradient from 1 to 51 $\text{pS}/\mu\text{m}^2$ at apical dendrites. Modeling details are referred to Table 1 and Table 2.

Table 1: Primate model parameter

	Apical dendrite	Basal dendrite	Soma	Axon
Length (μm)	1800	750	30	75
Diameter (μm)	9	$0.375N$	15	1.5
R_{axial} ($\Omega \cdot \text{cm}$)	150	150	150	50
Capacitance ($\mu\text{F}/\text{cm}^2$)	1	1	1	1
R_{leak} ($\Omega \cdot \text{cm}^2$)	$10000/(1 - 0.8x)$	14000	14000	14000
g_{Na} (S/cm^2)	0.01	0.002	0.04	0.8
$g_{\text{K}_{\text{ap}}}$ (S/cm^2)	$0.0001(1 + 50x)$	0.0001	0.01	0.4
$g_{\text{K}_{\text{dr}}}$ (S/cm^2)	0.01	0.01	0.01	0.4
E_{leak} (mV)	-65	-65	-65	-65
E_{Na} (mV)	55	55	55	55
$E_{\text{K}_{\text{ap}}}$ (mV)	-90	-90	-90	-90
$E_{\text{K}_{\text{dr}}}$ (mV)	-90	-90	-90	-90

Table 2: Mouse model parameter

	Apical dendrite	Basal dendrite	Soma	Axon
Length (μm)	900	500	20	50
Diameter (μm)	6	$0.25N$	10	1
R_{axial} ($\Omega \cdot \text{cm}$)	150	150	150	50
Capacitance ($\mu\text{F}/\text{cm}^2$)	1	1	1	1
R_{leak} ($\Omega \cdot \text{cm}^2$)	$35000/(1 + 24x^3)$	40000	40000	40000
g_{Na} (S/cm^2)	0.01	0.002	0.04	0.8
$g_{\text{K}_{\text{ap}}}$ (S/cm^2)	$0.0001(1 + 50x)$	0.0001	0.01	0.3
$g_{\text{K}_{\text{dr}}}$ (S/cm^2)	0.01	0.01	0.01	0.3
E_{leak} (mV)	-65	-65	-65	-65
E_{Na} (mV)	55	55	55	55
$E_{\text{K}_{\text{ap}}}$ (mV)	-90	-90	-90	-90
$E_{\text{K}_{\text{dr}}}$ (mV)	-90	-90	-90	-90

Note. N stands for basal dendrite number, and x is normalized distance from soma ($x = 0$) to distant apical tip ($x = 1$).

C Reliability analysis

To test the response reliability of neurons with various dendrites, we inject a current following an Ornstein-Uhlenbeck (OU) process. The τ_{OU} in OU process is 5 ms. In reliability tests, one of the OU currents is applied at the middle point of soma to simulate neuronal-specific background inputs, the other is applied at the middle point of a basal dendrite (basal dendrites are symmetrical) with doubled current density to simulate a stronger and shared feedforward inputs across neurons. Current density is tuned to generate firing rate of 15 Hz for neuron with two dendrites while the density ratio holds, as shown in Fig. S2(a). The density is fixed for other neurons with various dendrites.

Spike timing reliability was assessed by calculating the correlation coefficient of spike trains across repeated stimulus presentations. For a set of repeated trials, individual spike trains were discretized into binary vectors $s_i(t)$ and $s_j(t)$ (where i, j denote trial indices), with $s(t) = 1$ indicating a spike event. Spikes are identified by 2 ms time bins, within which the membrane potential exceeds 0 mV. We calculated the pairwise cross-covariance at zero lag (C_{ij}) for all unique pairs where $i \neq j$. The reliability coefficient R was derived as the mean of these zero-lag cross-covariances:

$$R = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{C_{ij}(0)}{\sqrt{C_{ii}(0) C_{jj}(0)}}. \tag{3}$$

where N is the total number of trials. This measure normalizes the reliability score between 0 (uncorrelated noise) and 1 (perfect temporal reproducibility). We perturb the τ_{OU} in dendritic OU current from 3 ms to 10 ms (Fig. S2(b)), yielding similar results to the reported 5 ms in Fig. 3. Notably, we study the morphology contribution by mouse neurons with expanded size as human neurons. The reliability of those human-sized mouse neurons slightly deviates from typical mouse neurons (Fig. S2(b), grey dash line), indicating that the resistance redistribution of human neurons plays dominant role between the reliability gap.

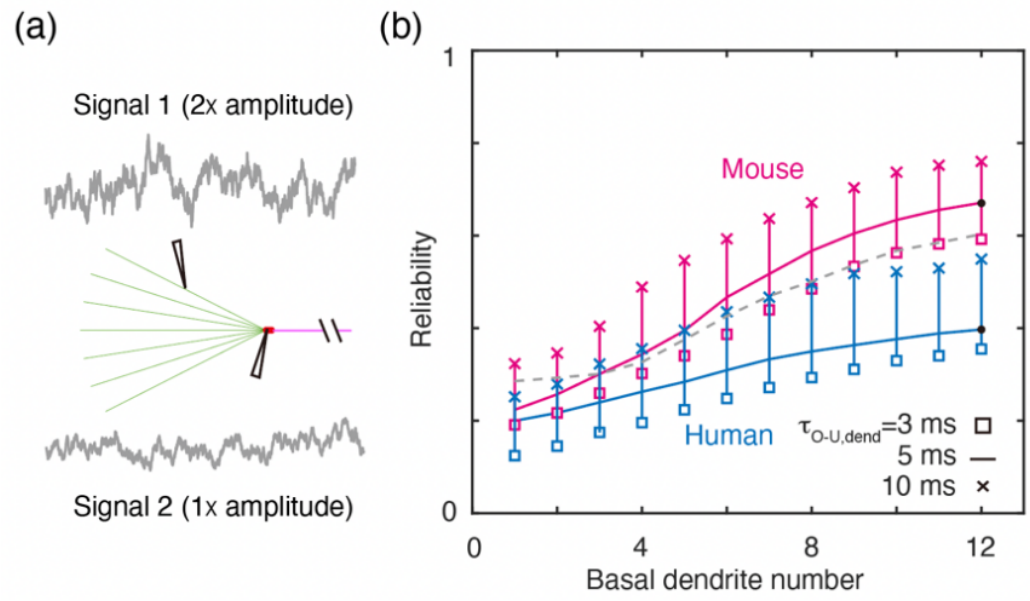


Figure S2: (a) Stimuli protocol of dual O-U currents. (b) Measured reliability of mouse and human neurons with different τ_{OU} in dendritic OU process. The grey dash line denotes the reliability of mouse neurons with expanded size as human neurons.

D Flexibility analysis

To test flexibility as the coefficient of variance of inter-spike interval (CV_{isi}), only one OU current on soma is present, and the current density varies to cover a wide range of firing rate (2–10 Hz).

Spikes are identified when the membrane potential exceeds 0 mV in a 2 ms time bin. In Fig. 3, results correspond to neuron firing rate at 2 Hz, each point is averaged from 50 neurons, and the activity of each neuron is simulated for 10 seconds. CV_{isi} at other fire rate are close to the one at 2 Hz, as presented in Fig. S3 (a) and (b). Notably, we also perform test on mouse neurons with human-derived morphology, as shown in Fig. S3 (c). The size-expanded mouse neuron has identical membrane time constant as the typical mouse neuron. We find that the expanded size introduced limited flexibility, that neurons with more than 12 dendrites begin to burst. Nevertheless, human neurons with redistributed biophysiological properties have prominently greater flexibility.

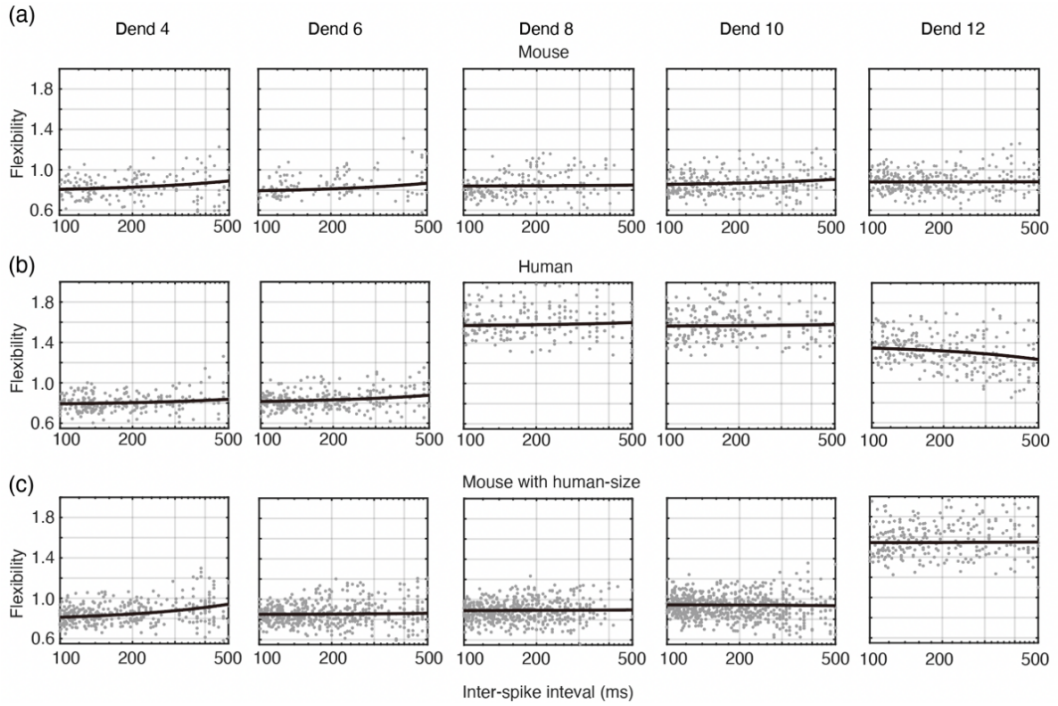


Figure S3: The coefficient of variation from 4-dendrite neurons to 12-dendrite neurons, in step of 2 dendrites. The mouse neurons (a) spike regularly with slight flexibility increase with basal dendrite count, and human neurons (b) are capable of bursting with dendrites more than 8. Mouse neurons with expanded size as human neurons (c) are capable of bursting with dendrites more than 12.

E Mutual information (MI) and coding efficiency (CE) analysis

We quantify the information transmission amount of the neuron using the "direct method" as described by Strong et al. [51]. This approach calculates the mutual information (I) between the stimulus and the neural response as the difference between the total entropy (S) of the spike train and the noise entropy (N):

$$I = S - N. \tag{4}$$

Neural responses are first discretized into binary spike trains (1 for spike, 0 for no spike) with a temporal resolution of dt ms. We defined a neural "word" as a binary sequence of length L bins. Using a sliding window approach, the binary pattern within each window $[t, t + L - 1]$ is converted into a decimal integer, creating a matrix of response words $W(i, t)$, where i denotes the word index and t denotes the time bin index.

E.1 Total Entropy (S)

The total entropy refers to the total amount of information that can be encoded in the spike train. To calculate S, we pooled the word responses from all trials and all time points into a single distribution.

The probability of occurrence for each unique word w, denoted as $P(w)$, was estimated from this aggregated dataset. The total entropy then was calculated as:

$$S = -\sum_{w} P(w) \log_2 P(w), \tag{5}$$

which is the entropy of the marginal distribution of words.

E.2 Noise Entropy (N)

The noise entropy quantifies the variability of the neural response to repeated presentations of the same stimulus and reflects response variability unrelated to stimulus information. We calculated the local entropy at each specific time point t across all trials. For a fixed time t, we estimated the probability distribution of words $P(w|t)$ across the repeated trials. The local noise entropy $H(t)$ was defined as:

$$H(t) = -\sum_{w} P(w|t) \log_2 P(w|t). \tag{6}$$

The global noise entropy N was obtained by averaging these local entropy over all time points in the stimulus duration:

$$N = \langle H(t) \rangle_t . \tag{7}$$

E.3 Mutual Information

The mutual information was computed by subtracting the noise entropy from the total entropy ($I = S - N$), yielding the information rate in bits per word.

In our implementations, we set dt as 1 ms, and words as 8 bins, similar to that in previous work [47].

E.4 Coding efficiency

To quantify the effectiveness of the neural code, we computed the coding efficiency (CE). This metric represents the fraction of the system's total capacity (total entropy) that is utilized to transmit stimulus-specific information. It was calculated as the ratio of the mutual information to the total entropy:

$$CE = \frac{I}{S}. \tag{8}$$

Note that this value is dimensionless as a percentage, and provides a normalized measure of information transmission independent of the firing rate. We conclude total entropy, mutual information and noise entropy in Fig. S4. We perturbate the τ_{OU} in dendritic OU current from 3 ms to 10 ms (same as Fig. S2), yielding similar results to the reported 5 ms in Fig. 3.

F Analytical solution of neuronal time constant with multiple basal dendrites

To derive the effective time constant (τ_{eff}) of a neuron, we employ the admittance analysis method from linear systems theory. This approach accurately captures the charge distribution delays caused by the physical structure of dendrites, specifically the impact of axial resistance.

F.1 Introduction of the Complex Frequency Operator

In the time domain, neuronal dynamics are described by differential equations. To transform these differential operations into simpler algebraic equations, we apply the Laplace transform, mapping the time variable t to the complex frequency variable s. Under this framework, the time-derivative

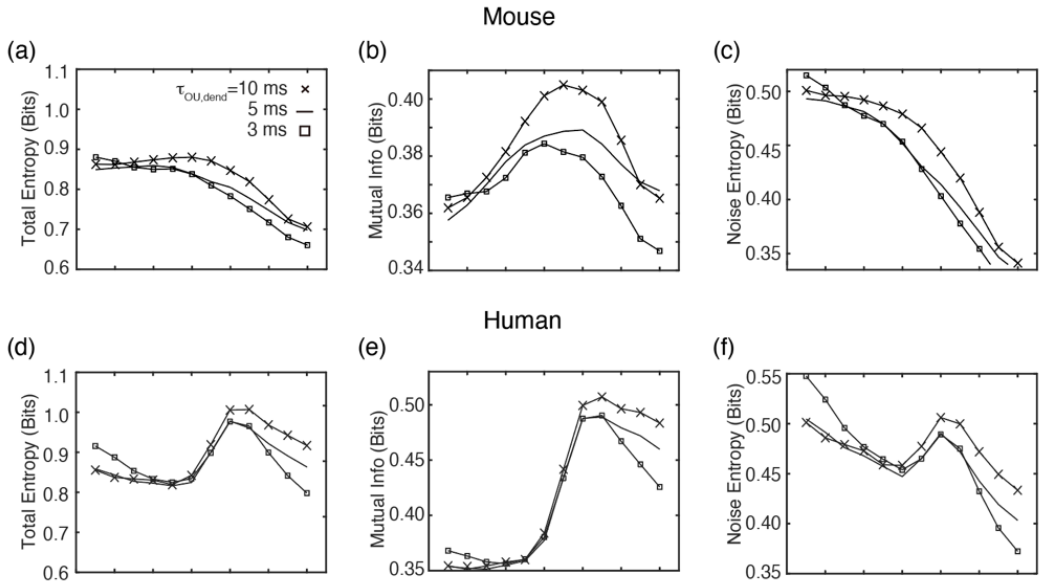


Figure S4: (a-c) Information entropy, mutual information and noise entropy of mouse neurons, and (d-f) same for human neurons, based on the reliability tests with two O-U inputs. Line with crosses denote dendrite O-U input with $\tau_{\mathrm{O-U,dend}} = 10$ ms, and 5 ms for bare line and 3 ms for line with squares.

operator d/dt corresponds to the multiplication operator s. This allows us to define Admittance (Y), which is the ratio of current to voltage in the frequency domain:

$$Y(s) = \frac{I(s)}{V(s)}. \tag{9}$$

F.2 Dynamical Equations for Single Basal Dendrite

Consider a single dendritic compartment connected to the soma via an axial resistance r_a. Based on Kirchhoff's Current Law (KCL) and Ohm's Law, we first consider axial current equation. The current I_a flowing from the soma to the dendrite is driven by the potential difference between the somatic voltage $V_{\mathrm{soma}}(s)$ and the dendritic voltage $V_{\mathrm{dend}}(s)$:

$$I_a(s) = \frac{V_{\mathrm{soma}}(s) - V_{\mathrm{dend}}(s)}{r_a}. \tag{10}$$

From the perspective of dendritic membrane balance, the current injected into the dendrite must equal to the sum of the capacitive displacement current and the ionic leak current:

$$I_a(s) = sC_{\mathrm{dend}}V_{\mathrm{dend}}(s) + g_{\mathrm{dend}}V_{\mathrm{dend}}(s) = V_{\mathrm{dend}}(s)\left(g_{\mathrm{dend}} + sC_{\mathrm{dend}}\right). \tag{11}$$

where g_{dend} represents the dendritic membrane conductance, C_{dend} is the dendritic membrane capacitance; and the term sC_{dend} represents the capacitive component of the current in the frequency domain. Combining the equations,

$$V_{\mathrm{dend}}(s) = \frac{I_a(s)}{g_{\mathrm{dend}} + sC_{\mathrm{dend}}}. \tag{12}$$

Substituting this into the axial current equation, we have:

$$I_a(s)r_a = V_{\text{soma}}(s) - \frac{I_a(s)}{g_{\text{dend}} + sC_{\text{dend}}}. \tag{13}$$

Rearranging the equation, we obtain the admittance of a single basal dendrite:

$$Y_{\text{dend}}(s) = \frac{I_a(s)}{V_{\text{soma}}(s)} = \frac{g_{\text{dend}} + sC_{\text{dend}}}{1 + r_a\left(g_{\text{dend}} + sC_{\text{dend}}\right)}. \tag{14}$$

A neuron with N identical parallel dendrites connected to a soma has a total input current I_{in} equal to the somatic transmembrane current plus the sum of all axial currents:

$$Y_{\text{total}}(s) = \frac{I_{\text{in}}(s)}{V_{\text{soma}}(s)} = (g_{\text{soma}} + sC_{\text{soma}}) + NY_{\text{dend}}(s). \tag{15}$$

where g_{soma} and C_{soma} are the membrane conductance and capacitance of the soma, respectively.

Note that the effective time constant τ_{eff} describes the system's decay characteristics as it approaches a steady state. This corresponds to a linear approximation of $Y_{\text{total}}(s)$ near $s \to 0$.

To the effective conductance G_{eff}:

$$G_{\text{eff}} = Y_{\text{total}}(0) = g_{\text{soma}} + \frac{Ng_{\text{dend}}}{1 + r_a g_{\text{dend}}}. \tag{16}$$

To the effective capacitance C_{eff}:

$$C_{\text{eff}} = \left.\frac{dY_{\text{total}}(s)}{ds}\right|_{s\to 0} = C_{\text{soma}} + N\frac{d}{ds}\left[\frac{g_{\text{dend}} + sC_{\text{dend}}}{1 + r_a\left(g_{\text{dend}} + sC_{\text{dend}}\right)}\right]. \tag{17}$$

in which,

$$\frac{d}{ds}\left(\frac{g_{\text{dend}} + sC_{\text{dend}}}{1 + r_a\left(g_{\text{dend}} + sC_{\text{dend}}\right)}\right) = \frac{C_{\text{dend}}\left[1 + r_a\left(g_{\text{dend}} + sC_{\text{dend}}\right)\right] - r_a C_{\text{dend}}\left(g_{\text{dend}} + sC_{\text{dend}}\right)}{\left[1 + r_a\left(g_{\text{dend}} + sC_{\text{dend}}\right)\right]^2}.$$

When $s = 0$,

$$\frac{d}{ds}\left(\frac{g_{\text{dend}} + sC_{\text{dend}}}{1 + r_a\left(g_{\text{dend}} + sC_{\text{dend}}\right)}\right) = \frac{C_{\text{dend}}}{\left(1 + r_a g_{\text{dend}}\right)^2}.$$

In brief,

$$C_{\text{eff}} = C_{\text{soma}} + N\frac{C_{\text{dend}}}{\left(1 + r_a g_{\text{dend}}\right)^2}. \tag{18}$$

The squared term in the denominator indicates that the contribution of basal dendritic capacitance to the somatic potential is attenuated by the "electrotonic distance" created by the axial resistance r_a.

In conclusion, the system's effective time constant is defined as the ratio of effective capacitance to effective conductance:

$$\tau_{\text{eff}} = \frac{C_{\text{eff}}}{G_{\text{eff}}} = \frac{C_{\text{soma}} + N\frac{C_{\text{dend}}}{(1 + r_a g_{\text{dend}})^2}}{g_{\text{soma}} + \frac{Ng_{\text{dend}}}{1 + r_a g_{\text{dend}}}}. \tag{19}$$

F.3 Asymptotic analysis

As the number of dendritic branches N approaches infinity,

$$\tau_{\text{limit}} = \lim_{N\to\infty} \tau_{\text{eff}}(N) = \frac{C_{\text{dend}}}{g_{\text{dend}}\,(1 + r_a g_{\text{dend}})}. \tag{20}$$

By defining the intrinsic dendritic time constant $\tau_{\text{dend}} = C_{\text{dend}}/g_{\text{dend}}$, the asymptotic limit is:

$$\tau_{\text{limit}} = \frac{\tau_{\text{dend}}}{1 + r_a g_{\text{dend}}} = \frac{\tau_{\text{dend}}}{1 + g_{\text{dend}}/g_a}. \tag{21}$$

In typical neurons, $g_{\text{dend}} \ll g_a$, so $\tau_{\text{limit}} \approx \tau_{\text{dend}}$. On the other hand, As the number of dendritic branches N is 0, $\tau_{\text{eff}} = C_{\text{soma}}/g_{\text{soma}} = \tau_{\text{soma}}$. This explicitly explains the observation in Fig. 3. This result demonstrates that in the rich-dendrite limit, the neuron's time constant is independent of the soma and converges to a value close to the intrinsic dendritic time constant. Strictly speaking, the neuron's time constant is smaller than dendritic time constant, because the axial resistance limits the efficiency with which the soma charges distal dendritic compartments, accelerating the observed decay rate at the soma. In sparse-dendrite neurons, the neuron's time constant is dominated by somatic time biophysiological properties.

G Spiking network experiment for hierarchical time-constant heterogeneity

The main text identifies a dissociation between the mean and variance of effective neuronal time constants across cortical hierarchy. The mean effective time constant remains approximately matched across cortical areas, whereas time-constant variability follows different species-specific organizations. In primate/human-like neurons, variability decreases along the cortical hierarchy, while in mouse-like neurons it remains comparatively uniform across areas. To test whether this organization has a functional consequence for temporal computation, we construct a controlled spiking neural network experiment in which the mean time constant is matched across models, while the allocation of time-constant heterogeneity across layers is systematically varied.

The goal of this experiment is not to reproduce cortical anatomy in detail, but to isolate the computational role of hierarchical time-constant variance. We therefore use the same network architecture, training objective, dataset, optimizer, and nominal mean time constant across all model variants. The only experimental manipulation is the layer-wise profile of PLIF time-constant heterogeneity. This design allows us to ask whether a primate-like contraction of time-constant variability from early to late processing stages improves robustness to temporally structured distribution shifts, including changes in motion speed, luminance flicker, sensory noise, and their combinations.

G.1 Moving MNIST video classification task

We train and evaluate the networks on a single-digit Moving MNIST video classification task. Each sample is a grayscale video sequence

$$X = \{x_t\}_{t=1}^{T}, \qquad x_t \in \mathbb{R}^{1\times H\times W},$$

with default temporal length $T = 16$ and spatial size $H = W = 64$. The class label is the identity of the moving digit,

$$y \in \{0, 1, \ldots, 9\},$$

which is the same ten-way label space as standard MNIST. Unlike static MNIST classification, the model must infer digit identity from a sparse, degraded, and temporally varying visual stream.

Each video is generated by randomly sampling one 28×28 MNIST digit and placing it on a 64×64 blank canvas. The digit moves linearly with constant speed and direction, and reflects elastically at the image boundaries. Let $p_t = (p_{x,t}, p_{y,t})$ denote the upper-left position of the digit bounding box at time t. The initial position p_1 is sampled uniformly from all locations for which the full digit lies inside the canvas. The initial velocity is

$$v = s(\cos\theta, \sin\theta),$$

where s is the speed in pixels per frame and θ is sampled uniformly from $[0, 2\pi)$. At each subsequent frame, the position is updated by

$$p_{t+1} = p_t + v.$$

If the digit crosses a horizontal or vertical boundary, the corresponding velocity component is reflected, so that the digit remains inside the canvas. This produces continuous motion trajectories with boundary reflections.

To increase the temporal demands of the task, each video contains several nuisance factors that are independent of digit identity. These factors include random initial position, motion speed, motion direction, frame-wise luminance flicker, and additive Gaussian noise. The flicker factor modulates the global brightness of the digit across time, so that the same digit appears with temporally fluctuating intensity. In implementation, the clean digit image at frame t is multiplied by a scalar flicker factor a_t, whose temporal frequency is controlled by the sampled flicker frequency. If a sinusoidal modulation is used, this can be written as

$$a_t = 1 + A_{\text{flicker}} \sin(2\pi f_{\text{flicker}} t + \phi),$$

where A_{flicker} is the flicker amplitude, f_{flicker} is the flicker frequency, and ϕ is a random phase. In the implementation, the exact flicker amplitude and clipping range are specified in the configuration file. The generated frame is clipped to the valid image range after flicker and noise are applied.

We further impose sparse foreground corruption to make the task rely on temporal integration rather than single-frame recognition. Within the digit bounding box, foreground pixels are retained independently with probability

$$p_{\text{retain}} = \texttt{sparse_fg_retain_frac},$$

with default value

$$p_{\text{retain}} = 0.25.$$

All remaining foreground pixels are set to zero. This operation produces a sparse and fragmented digit appearance in each frame. Because only a subset of the digit is visible at any moment, the network must aggregate evidence across time to classify the digit reliably. Unless otherwise specified, this sparse foreground mask is sampled independently for each video sample and frame.

After sparse foreground corruption and flicker modulation, additive Gaussian noise can be applied to each frame. The image is clipped to the valid intensity range after noise injection. This noise manipulation is used both during training and evaluation, depending on the split.

G.2 Training and evaluation distributions

The training distribution contains a finite set of motion speeds, flicker frequencies, and noise levels. Out-of-distribution evaluation is performed by holding out specific temporal factors from training and testing whether the learned representation generalizes to unseen temporal regimes. This setting is designed to test temporal robustness rather than memorization of a fixed motion or flicker distribution.

The training speed set is

$$\mathcal{S}_{\text{train}} = \{1, 2, 3, 4, 8\},$$

measured in pixels per frame. The held-out OOD speed is

$$s_{\text{OOD}} = 6,$$

which is not present during training. Therefore, the `test_ood_speed` and `test_combined_ood` splits evaluate generalization to an unseen motion speed rather than standard in-distribution recognition.

The training flicker set contains one no-flicker condition and five nonzero temporal frequencies:

$$\mathcal{F}_{\text{train}} = \left\{ \text{none}, \frac{1}{16}, \frac{1}{12}, \frac{1}{8}, \frac{1}{6}, \frac{1}{4} \right\}.$$

For the flicker conditions, frame intensity is multiplicatively modulated as

$$x_t \leftarrow x_t \left(0.75 + 0.25 \sin(2\pi f t + \phi)\right),$$

Table 3: **Moving MNIST data-generation parameters.** These values correspond to the main Moving MNIST experiments. The separate sanity configuration includes an additional training noise level of 0.06.

Parameter	Value	Description
Video length	$T = 16$	Number of frames per sample
Canvas size	64×64	Spatial size of each frame
Digit size	28×28	MNIST digit image size
Number of classes	10	Digit identity classes
Training speeds	$\{1, 2, 3, 4, 8\}$	Pixels per frame
OOD speed	6	Held-out speed
Training flicker set	$\{\text{none}, 1/16, 1/12, 1/8, 1/6, 1/4\}$	Frequencies in cycles per frame
OOD flicker	$1/32$	Held-out flicker frequency
Flicker modulation	$0.75 + 0.25 \sin(2\pi f t + \phi)$	Multiplicative brightness modulation
Flicker amplitude	0.25	Sinusoidal coefficient; scale range approximately $[0.5, 1.0]$
Sparse foreground retention	0.25	Probability of retaining foreground pixels
Training noise levels	$\{0.0, 0.03, 0.10\}$	Pixel-wise Gaussian noise standard deviations
ID noise std	0.0	In-distribution test noise level
OOD noise std	0.06	Held-out noise level for OOD noise evaluation

where f is measured in cycles per frame and ϕ is a random phase. Thus, flicker scales the digit intensity between approximately 0.5 and 1.0 of its original value. The held-out OOD flicker frequency is

$$f_{\text{OOD}} = \frac{1}{32},$$

which is not included in the training flicker set. Therefore, the `test_ood_flicker` and `test_combined_ood` splits evaluate generalization to an unseen luminance fluctuation timescale.

Training noise levels are sampled from

$$\Sigma_{\text{train}} = \{0.0, 0.03, 0.10\},$$

where each value denotes the standard deviation of additive pixel-wise Gaussian noise applied before clipping the frame to the valid intensity range. Noise OOD is controlled separately from speed and flicker OOD. For example, the in-distribution test set uses

$$\sigma_{\text{noise}} = 0.0,$$

whereas the OOD noise split uses

$$\sigma_{\text{noise}} = 0.06,$$

which is not included in the main training noise set. This design separates OOD noise from OOD discrete temporal factors. The combined OOD split evaluates simultaneous shifts in speed, flicker frequency, and noise level.

G.3 Network architecture

We use a compact convolutional spiking neural network, denoted `MovingMNISTSNN`. The network contains four convolutional processing stages. Each stage consists of a two-dimensional convolution, batch normalization, a profile-controlled PLIF neuron, and spatial average pooling. All convolutional layers use 3×3 kernels with padding 1 and no convolutional bias. Batch normalization follows each convolution. A profile-controlled PLIF node[18] then converts the normalized current into membrane dynamics and spikes. Spatial downsampling is performed by average pooling. After the final stage, global average pooling converts the last-layer spatial feature map into a vector representation.

For an input video of length T, the same convolutional spiking network is applied recurrently over time. At each time step, the frame x_t is passed through the convolutional hierarchy, while each PLIF neuron maintains its own membrane state. The hidden membrane potential and spike output therefore depend not only on the current frame, but also on previous input history through the membrane dynamics. All membrane states are reset between video samples.

G.4 Profile-controlled PLIF neurons

Each spiking unit is implemented as a parametric leaky integrate-and-fire neuron with a learnable membrane time constant. We use PLIF neurons because they allow the temporal integration scale to be optimized during training while preserving a biologically interpretable membrane variable [18]. In our experiments, the key manipulation is not whether the time constant is learnable, but how the variance of time constants is distributed across layers.

For a PLIF neuron with membrane potential V_t, input current I_t, reset potential V_{reset}, firing threshold V_{th}, and membrane time constant τ, the pre-spike membrane update can be written as

$$\tilde{V}_t = V_{t-1} + \frac{1}{\tau}\left(I_t - (V_{t-1} - V_{\mathrm{reset}})\right).$$

The spike output is

$$S_t = H(\tilde{V}_t - V_{\mathrm{th}}),$$

where $H(\cdot)$ is the Heaviside step function. After spike generation, the membrane potential is reset according to

$$V_t = \tilde{V}_t(1 - S_t) + V_{\mathrm{reset}} S_t.$$

We use threshold

$$V_{\mathrm{th}} = 1.0$$

and reset potential

$$V_{\mathrm{reset}} = 0.0.$$

The reset operation is detached from the computational graph, corresponding to `detach_reset=True`. Gradients through the spike function are estimated using the arctangent surrogate gradient, `surrogate.ATan()`.

The time constant is parameterized separately for each layer and channel. Let $u_{\ell,c}$ be the learnable raw time-constant variable for channel c in layer ℓ. Within each layer, the raw vector is normalized to have zero mean and unit variance:

$$\tilde{u}_{\ell,c} = \frac{u_{\ell,c} - \mu_\ell}{\max(\sigma_\ell, \sigma_{\mathrm{floor}})},$$

where

$$\mu_\ell = \frac{1}{C_\ell}\sum_{c=1}^{C_\ell} u_{\ell,c}, \qquad \sigma_\ell = \sqrt{\frac{1}{C_\ell}\sum_{c=1}^{C_\ell}(u_{\ell,c} - \mu_\ell)^2},$$

and σ_{floor} is the variance floor specified by `tau_u_variance_floor`. The time constant is then constructed as

$$\tau_{\ell,c} = \mathrm{clip}\left(\tau_0 \exp(s_\ell \tilde{u}_{\ell,c}), \tau_{\min}, \tau_{\max}\right),$$

where τ_0 is the nominal time constant, s_ℓ is the layer-specific profile scale, and $\tau_{\min}$ and $\tau_{\max}$ define the allowed time-constant range. This construction keeps the nominal mean timescale approximately matched while allowing the variance of τ to be controlled by s_ℓ.

The profile scale s_ℓ is the only difference among the main experimental model variants. Larger s_ℓ produces larger channel-wise time-constant diversity in layer ℓ, whereas $s_\ell = 0$ collapses all channels toward the nominal value τ_0 up to clipping. Because the normalized variable $\tilde{u}_{\ell,c}$ has approximately zero mean and unit variance, the layer-wise variance of $\log \tau_{\ell,c}$ is primarily controlled by s_ℓ^2 before clipping.

G.5 Time-constant heterogeneity profiles

We compare four time-constant heterogeneity profiles. The *primate* profile assigns high time-constant variance to early layers and progressively lower variance to deeper layers:

$$s_{\mathrm{primate}} = [0.80, 0.55, 0.35, 0.15].$$

This profile mirrors the primate/human-like pattern observed in the biological analysis, where time-constant variability is larger in early cortical areas and smaller in higher-order cortical areas.

Table 4: **Layer-wise time-constant heterogeneity profiles.** The primate, reverse, and mouse profiles have matched total variance budget, whereas the fixed profile removes meaningful heterogeneity.

Profile	Layer 1	Layer 2	Layer 3	Layer 4
Primate	0.80	0.55	0.35	0.15
Reverse	0.15	0.35	0.55	0.80
Mouse	0.522	0.522	0.522	0.522
Fixed	0.00	0.00	0.00	0.00

The *reverse* profile uses the opposite allocation:

$$s_{\text{reverse}} = [0.15, 0.35, 0.55, 0.80].$$

This condition tests whether the benefit depends specifically on assigning temporal diversity to early layers, rather than merely having the same amount of heterogeneity somewhere in the network.

The *mouse* profile assigns approximately equal variance to all layers:

$$s_{\text{mouse}} = [0.522, 0.522, 0.522, 0.522].$$

The value 0.522 is chosen so that the total variance budget approximately matches that of the primate and reverse profiles. Specifically,

$$\sqrt{\frac{0.80^2 + 0.55^2 + 0.35^2 + 0.15^2}{4}} \approx 0.522.$$

Therefore, the primate, reverse, and mouse profiles have comparable total heterogeneity strength, but differ in how this heterogeneity is distributed across depth.

The ***fixed*** profile removes meaningful time-constant heterogeneity:

$$s_{\text{fixed}} = [0, 0, 0, 0].$$

In this condition,

$$\tau_{\ell,c} \approx \tau_0$$

for all layers and channels, subject only to clipping by $[\tau_{\min}, \tau_{\max}]$. This condition serves as a low-heterogeneity control.

This profile design is central to the experiment. The primate, reverse, and mouse models are not compared because one has more total time-constant diversity than another. Instead, they are compared because they allocate a matched diversity budget differently across the hierarchy. The fixed model asks whether time-constant diversity itself is useful, whereas the primate, reverse, and mouse models ask where in the hierarchy this diversity should be placed.

G.6 Supervised contrastive training

The main representation-learning experiment uses supervised contrastive learning. The final classifier of the spiking network, `snn.fc`, is frozen during self-supervised pretraining. A projection head is attached to the spike-based representation produced by `forward_spike_repr`. The projection head is a two-layer multilayer perceptron:

$$g(z) = W_2 \,\mathrm{ReLU}(W_1 z + b_1) + b_2.$$

The projected representation is used for supervised contrastive loss, while an auxiliary online linear classifier is trained on a detached version of the spike representation.

The spike representation is computed from the last-layer spikes. Let $S_t^{(4)}$ denote the spike tensor from the fourth PLIF layer at time t. We first apply global average pooling to obtain a vector representation for each selected time step:

$$r_t = \mathrm{GAP}(S_t^{(4)}).$$

The final representation averages only the last K time steps:

$$z = \frac{1}{K} \sum_{t=T-K+1}^{T} r_t.$$

The parameter K is specified by `last_k_steps`. This representation emphasizes the late dynamic state of the network after temporal evidence has accumulated across the video.

For supervised contrastive learning, each batch is class-balanced. The batch size must satisfy

$$B = 10 \times N_{\text{per class}},$$

where $N_{\text{per class}}$ is the number of samples per digit class in the batch. A stratified batch sampler enforces this constraint. This design ensures that each anchor has positive examples from the same class within the batch, which is required for stable supervised contrastive training.

Let q_i be the normalized projection of sample i:

$$q_i = \frac{g(z_i)}{\|g(z_i)\|_2}.$$

For an anchor i, let $P(i)$ denote the set of other samples in the batch with the same class label. The supervised contrastive loss is

$$\mathcal{L}_{\text{supcon}} = -\sum_{i=1}^{B} \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(q_i^\top q_p / T_{\text{con}})}{\sum_{a \neq i} \exp(q_i^\top q_a / T_{\text{con}})},$$

where T_{con} is the contrastive temperature specified by `temperature`. This objective encourages representations of the same digit identity to cluster together while separating different digit identities, despite temporal nuisance variation.

An auxiliary online classifier is trained on the detached spike representation:

$$\hat{y}_{\text{online}} = W_{\text{online}} \,\text{stopgrad}(z) + b_{\text{online}}.$$

The online classifier uses cross-entropy loss,

$$\mathcal{L}_{\text{online}} = \text{CE}(\hat{y}_{\text{online}}, y).$$

Because z is detached, this loss updates only the online classifier and does not backpropagate into the spiking backbone. This prevents the classification head from providing an overly direct supervised shortcut to the backbone during contrastive pretraining. The total loss is

$$\mathcal{L} = \mathcal{L}_{\text{supcon}} + \lambda_{\text{online}} \mathcal{L}_{\text{online}},$$

where λ_{online} is specified by `online_ce_weight`.

The network is trained for `epochs_pretrain` epochs using a cosine learning-rate schedule. All model variants use the same training settings.

G.7 Linear-probe evaluation

After Supervised contrastive pretraining, we evaluate the quality and robustness of the learned spike representation using a frozen-backbone linear probe. The spiking network is frozen, and a linear classifier is trained on the representation z produced by `forward_spike_repr`. We refer to this evaluation model as `SpikeProbeNet`. The probe receives the same representation used by the contrastive objective, namely the average-pooled last-layer spike representation over the final K time steps.

The probe is trained on the training distribution and evaluated on multiple test splits. The primary metric is classification accuracy:

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}[\hat{y}_i = y_i].$$

We report accuracy separately for the in-distribution test set and for each OOD split. The OOD splits include held-out speed, held-out flicker frequency, held-out noise level, and combined OOD perturbations. The combined OOD split is the most stringent condition because it simultaneously changes motion speed, luminance fluctuation, and noise relative to training.

The online classifier trained during pretraining can also be used as an auxiliary probe. However, the main evaluation uses the frozen-backbone linear probe, because it provides a consistent post hoc measurement of representation quality across all time-constant profiles.

Table 5: **Evaluation splits.** Each split tests a distinct form of temporal or sensory distribution shift.

Split	Speed	Flicker frequency	Noise std
`test_id`	From training set	From training set	Typically 0.0
`test_ood_speed`	6	From training set	As configured
`test_ood_flicker`	From training set	$1/32$	As configured
`test_ood_noise`	From training set	From training set	Typically 0.06
`test_combined_ood`	6	$1/32$	As configured

G.8 Controlled comparison across time-constant profiles

The experiment compares four model profiles under identical data, architecture, and optimization settings. The primate profile tests the biological hypothesis derived from the main analysis: early layers benefit from broad time-constant diversity because they must encode rapidly changing, noisy, and fragmented sensory evidence, whereas deeper layers benefit from reduced diversity because they must stabilize class-level representations. The reverse profile tests whether the opposite hierarchy is less effective. The mouse profile tests whether a uniform distribution of time-constant variance across depth is sufficient. The fixed profile tests whether removing time-constant diversity impairs temporal robustness.

This comparison is controlled in three ways. First, the nominal mean time constant τ_0 is the same across all models. Second, the primate, reverse, and mouse profiles have matched total variance budget, so performance differences cannot be explained by one model having more overall time-constant heterogeneity. Third, all models are trained with the same data splits, batch structure, optimization schedule, and evaluation protocol. Therefore, any systematic performance difference across profiles reflects the layer-wise allocation of time-constant diversity.

G.9 Implementation details

All videos are generated online or from a deterministic dataset object using fixed random seeds. The same seed list is used across model profiles to ensure that differences are not driven by different sampled videos or initializations. For each random seed, all four profiles are trained and evaluated under the same data-generation parameters. Membrane states are reset at the start of each video. During training, gradients are backpropagated through time over the full $T = 16$ frame sequence.

All input frames are represented as single-channel floating-point tensors. After digit placement, sparse foreground masking, flicker modulation, and noise addition, pixel values are clipped to the valid range. The input tensor is arranged as a time-indexed sequence with shape

$$[T, 1, H, W]$$

for each sample, or as a batched tensor with shape

$$[B, T, 1, H, W]$$

depending on the implementation. The network processes the sequence one frame at a time and accumulates PLIF membrane states internally.

The PLIF implementation uses arctangent surrogate gradients for spike backpropagation. The reset operation is detached from the gradient graph. This follows standard spiking neural network training practice and avoids unstable gradients through the hard reset. Batch normalization is applied within each convolutional stage before the PLIF nonlinearity. The convolutional layers do not use bias terms because batch normalization provides affine normalization.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the morpho-temporal decoupling claims and cross-species dendritic scaling, and these are supported by the results and simulations, with scope clarified in the Discussion and Limitations.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The Discussion includes a dedicated limitations paragraph covering simplified morphology, passive assumptions, and dataset constraints, which bound the conclusions.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The equations are used for modeling and derivations in simulations rather than formal theorems or proofs requiring stated assumptions.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The Methods and Appendix detail morphology sources, model parameters, stimuli, and analysis protocols (e.g., OU inputs and coding metrics), while the main text summarizes the core setup for the primary simulations.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The paper cites public morphology datasets, but it does not provide a code or data repository with reproduction instructions.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The Methods and Appendix specify model geometry, biophysical parameters, stimulation protocols, and analysis settings for the main simulations and coding metrics.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The morphology analyses report sample sizes and statistical tests with p-values (e.g., Dunn's test), and figures define variability measures such as s.e.m. or s.d. across trials or inputs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: Hardware specifications and runtime or compute budgets are not reported for the simulations or network training.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work relies on publicly available morphology datasets and simulations, with no new human or animal experiments, and thus conforms to standard ethical guidelines for use of public data.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

Justification: The Broader Impact section in included in 'Summary and Conclusions' and discusses conceptual benefits for neuroscience and neuromorphic design and notes limited direct societal risks.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release high-risk models or scraped datasets; it focuses on simulations and analysis of public morphology data.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [No]

Justification: Existing datasets and software are cited, but explicit license or terms-of-use statements are not provided in the paper or supplement.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The submission does not release new datasets, code, or model artifacts as standalone assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The study does not involve crowdsourcing or human-subject experiments.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No human-subject studies are conducted, so IRB approval does not apply.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not used as a core methodological component of the modeling or analyses described.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.