Manuscript

ar009 · 21 June 2026 · Draft

Abstract

Gamma oscillations are widespread in cortical activity but are largely absent from trained spiking neural networks, which typically operate in a current-based regime or impose oscillations as an external input. We train a spiking network with a fixed pyramidal–interneuron gamma (PING) loop on MNIST under surrogate-gradient descent, with the recurrent E↔I weights held at biophysical values. At matched test accuracy on the accuracy–rate frontier, the post-training excitatory firing rate is roughly an order of magnitude below a conductance-based control (7-fold by total population spike count once the higher-rate inhibitory pool is included), and the trained rate is well described by an affine relation $r_E \approx 1.15 + 0.183 f_\gamma$ with the measured gamma frequency ($R^2 = 0.994$). When the loop weights are released for training under a Dale's-law clamp, the rhythmicity collapses within one epoch and is not recovered from any initial condition tested. The trained network classifies a continuously concatenated digit stream without retraining or a segmentation cue; streaming accuracy is approximately governed by the product of presentation duration and input rate, remains below $\approx 80\%$ for durations of 15 ms or less, and at 200 ms rises from chance below 0.5 Hz to clearly informative performance by 2 Hz. These results are consistent with an interpretation of gamma as a structural constraint on excitatory firing rates that does not require learned tuning of the inhibitory connectivity.

1. Introduction

Gamma oscillations in the 30–80 Hz band have been associated with attention, binding, and gating in cortical activity [**1, 2, 3**], with the original visual-cortex observation reported by [**4**]. Two generation mechanisms are commonly distinguished, ING and PING [**5, 6**] the present work focuses on PING. PING arises from the dynamics of a recurrent excitatory–inhibitory (E↔I) loop [**7**]. Optogenetic activation of fast-spiking interneurons in intact cortical circuits drives gamma rhythms [**8, 9, 10, 11**]. Earlier in vitro recordings [**12**] and biophysical models [**13**] characterised interneuron-driven gamma in isolated inhibitory networks, and the synaptic mechanisms that synchronise the interneuron pool have been described [**14**].

PING has been studied extensively in biophysical [**15, 16, 17**] and neural-mass / mean-field models [**18, 19, 20, 21**], but these models are descriptive: they are not trained on a task. The parallel literature on trainable spiking neural networks uses surrogate-gradient descent for end-to-end optimisation [**22, 23, 24**] the resulting networks are typically current-based and non-rhythmic. The rhythmic variants either impose the oscillation as an external input [**25**] or obtain it as an emergent property of unconstrained surrogate-gradient training [**26**] in neither case is the rhythm carried by a fixed, biophysically-calibrated PING loop, and the present work is to our knowledge the first task-trained spiking network of that kind.

Cortical pyramidal cells fire at low rates (typically below 10 Hz) under strong recurrent input [**27**], and the cortical metabolic budget is dominated by excitatory spike generation, with inhibitory spikes substantially less costly per spike [**28, 29**]. The mechanism that constrains pyramidal firing rates under these conditions is not fully understood. We test the hypothesis that a fixed E↔I loop, by generating a gamma rhythm, constrains the post-training excitatory firing rate. A spiking network with a fixed PING loop is trained on MNIST and the post-training firing rate is compared with a non-rhythmic baseline at matched accuracy. In the architecture studied here the measured rhythm and firing rate are tightly linked, so rate and timing are synergistic descriptions of a single dynamics rather than independent codes [**30**].

The remainder of the paper is organised as follows. §2.1 describes the model; §2.2 characterises the gamma onset in theory and simulation; §2.3 reports the trained accuracy–rate frontier; §2.4 reports two experiments that test whether the firing-rate reduction is acquired during training; §2.5 reports the relationship between the gamma frequency and the post-training firing rate; §2.6 reports robustness of the post-training firing rate to spike perturbations and to integration timestep; and §2.7 reports a streaming-classification protocol.

2. Results

2.1 The model: COBA baseline and the PING loop

The network is a two-population conductance-based spiking model with a single hidden layer of N_E excitatory (E) and N_I inhibitory (I) leaky integrate-and-fire (LIF) units, driven by feedforward input weights W_{in} and read out by a non-spiking leaky-integrator layer over the excitatory population (§5.2). Recurrence is confined to one excitatory–inhibitory (E↔I) loop: E projects to I through the weight W^{EI} and I back to E through W^{IE}, with no E→E or I→I connection. Throughout the paper we compare two configurations of this single architecture: with the loop disabled ($W^{EI} = W^{IE} = 0$) it is a conductance-based (COBA) control in which input drives the excitatory population alone, and with the loop engaged it is the pyramidal–interneuron gamma (PING) configuration. The same network is trained on MNIST by surrogate-gradient descent from §2.3 onward (§5.4); in this section we characterise its free-running dynamics.

Free-running, the COBA configuration fires asynchronously: its power spectral density (PSD) has no peak in the gamma band, and the excitatory firing-rate–current (f–I) curve rises to $\approx$ 485 Hz under the strongest drive tested. Engaging the E→I→E loop instead produces synchronous inhibitory bursts and gamma-banded excitatory rasters with a PSD peak at $f_\gamma \approx 42$ Hz, and holds the excitatory firing rate approximately an order of magnitude below the COBA baseline across two decades of input drive (Figure 1).

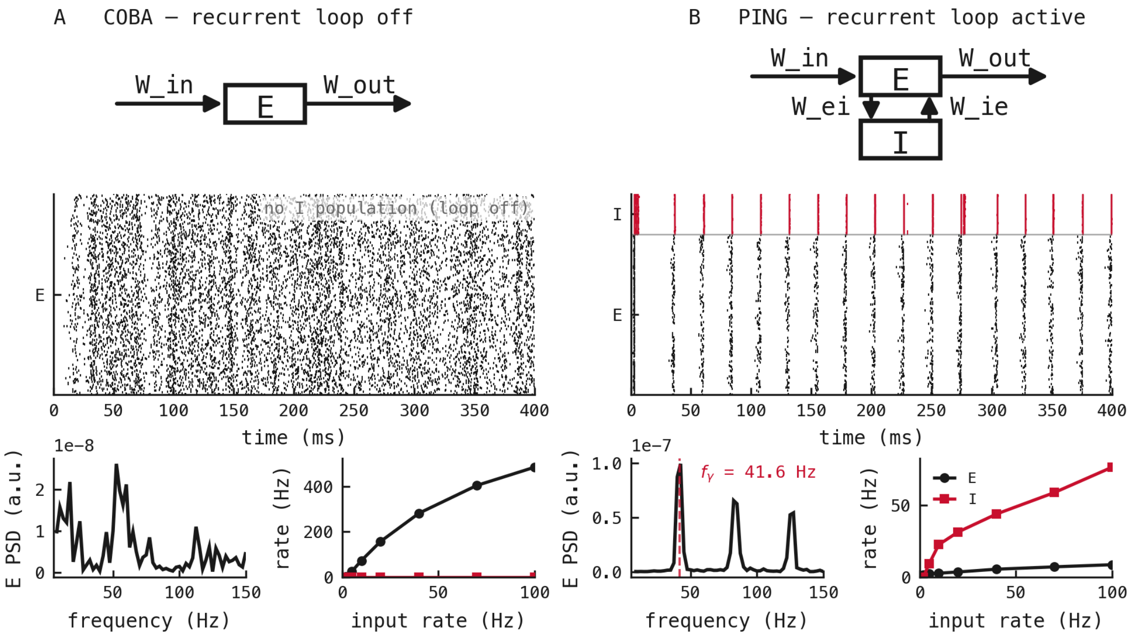


Figure 1: **A single recurrent E→I→E loop simultaneously generates a gamma rhythm and clamps the excitatory firing rate.** Free-running activity of the two-population conductance-based network (excitatory pool $N_E = 1024$, inhibitory pool $N_I = 256$; canonical biophysical parameters, §5.1) under matched Poisson drive, in two configurations. Each column shows, from top: a wiring schematic, a single-trial spike raster, the excitatory power spectral density (PSD), and the excitatory f–I curve. **(A)** COBA baseline with the recurrent loop disabled ($W^{EI} = W^{IE} = 0$): input projects to the excitatory (E) population only, with no inhibitory (I) population (schematic). The E spike raster is asynchronous, the Welch PSD of the summed E population shows no gamma-band peak, and the excitatory f–I curve rises to ≈ 485 Hz under the strongest drive tested. **(B)** PING configuration with the loop engaged (schematic: E→I via W^{EI}, I→E via W^{IE}; no I→I or E→E synapse): the inhibitory (I) population fires synchronous bursts, the E raster forms gamma bands, the PSD shows a discrete peak at $f_\gamma \approx 42$ Hz, and on axes shared with (A) the E rate is held approximately an order of magnitude lower across two decades of input drive while the I rate rises to ≈ 76 Hz. Source: exp023.

2.2 Gamma onset across the $W^{EI} \times W^{IE}$ plane

Across the $W^{EI} \times W^{IE}$ coupling plane the excitatory firing rate decreases, the inhibitory firing rate increases, and the lobe–trough rhythmicity score R (a normalised, dimensionless measure of periodicity in the population autocorrelation, bounded in $[0, 1)$; §5.5) increases monotonically with coupling strength: $R \approx 0$ along the COBA edges and $R \approx 0.98$ at strong coupling. The four-dimensional mean-field reduction (§5.3) predicts a supercritical Hopf bifurcation at external drive $I^\star_{\text{ext}} = 0.6$ nA with crossing frequency $f^\star \approx 27.8$ Hz. The classification as supercritical is supported by quasi-static up/down ramps with peak hysteresis below 10^{-5} in rate units and by the linear scaling of the squared steady-state oscillation amplitude A, $A^2 \propto (I_{\text{ext}} - I^\star_{\text{ext}})$ ($R^2 = 0.999$). The predicted gamma frequency is in qualitative agreement with the spiking measurement across the GABA synaptic decay time constant $\tau_{\text{GABA}} \in [4.5, 27]$ ms (Figure 2).

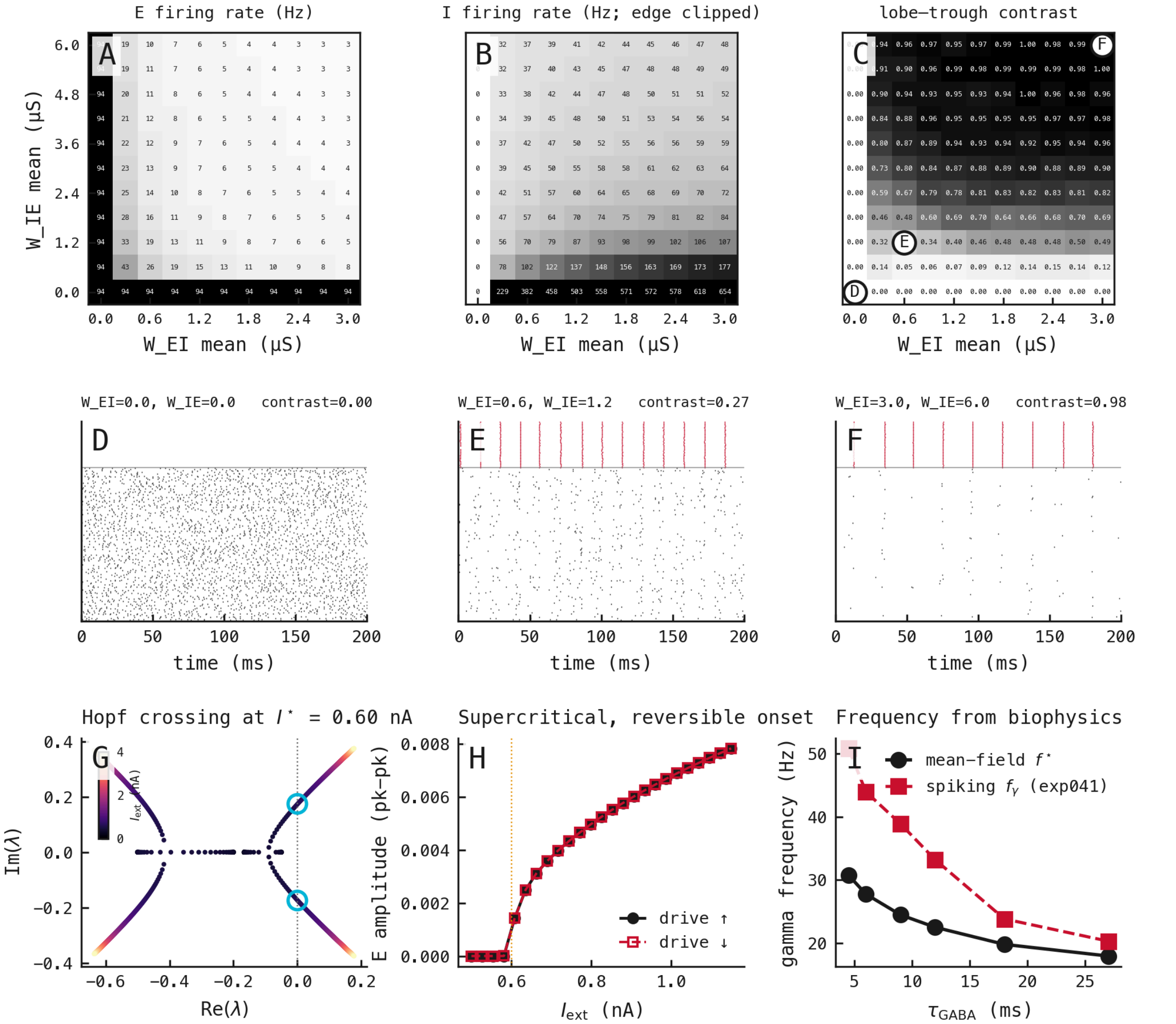


Figure 2: **Gamma emerges through a smooth, reversible onset across the coupling plane, consistent with a supercritical Hopf bifurcation of the mean-field reduction.** Nine panels (A–I). **(A–C)** Steady-state measurements across the 11×11 $W^{EI} \times W^{IE}$ coupling plane: mean excitatory firing rate (A), mean inhibitory firing rate (B), and the lobe–trough rhythmicity contrast R (C, §5.5), which is ≈ 0 along the two COBA edges and rises to ≈ 0.98 toward strong coupling. **(D–F)** Single-trial E (black) and I (red) rasters at three points sampled along the $W^{IE} = 2W^{EI}$ diagonal (circled in C): the loop-disabled origin (D, asynchronous), weak coupling ($R < 0.5$, E), and strong coupling (F, sharp gamma volleys). **(G–I)** The four-dimensional conductance mean-field reduction (§5.3): a complex-conjugate eigenvalue pair crosses into the right half-plane at external drive $I^{\star} = 0.6$ nA (G), locating a Hopf bifurcation at $f^{\star} \approx 27.8$ Hz; the steady-state oscillation amplitude grows continuously across the onset with coincident up- and down-ramp branches (H), the signature of a supercritical, reversible transition; and the predicted gamma frequency falls with τ_{GABA} in qualitative agreement with the spiking measurement (I). Source: exp054 (coupling-plane maps and mean-field, incorporating exp033); spiking f_γ from exp041.

2.3 Trained PING attains COBA accuracy at $\approx 10\times$ fewer spikes

Both architectures, trained on MNIST under surrogate-gradient descent (§5.4), converge to approximately 91% test accuracy. Sweeping the spike-budget penalty θ_u generates an accuracy–rate frontier; at every value of θ_u tested, PING attains higher accuracy at lower mean hidden-E firing rate than COBA. At the operating point with θ_u disabled, PING reaches $\approx 91\%$ accuracy at ≈ 12.3 Hz mean hidden-E rate, against $\approx 91\%$ at ≈ 181 Hz for COBA (Figure 3). The operating-point endpoints are reported as means over three seeds (42, 43, 44) and are approximately seed-invariant; interior points of the θ_u sweep that trace the frontier are run at a single seed (42). Seed-invariance of the post-training firing rate at finer resolution is established in §2.5–§2.7, where three seeds per condition are reported throughout. The PING rate does not decrease further when θ_u is lowered, consistent with a structural lower bound on the rate.

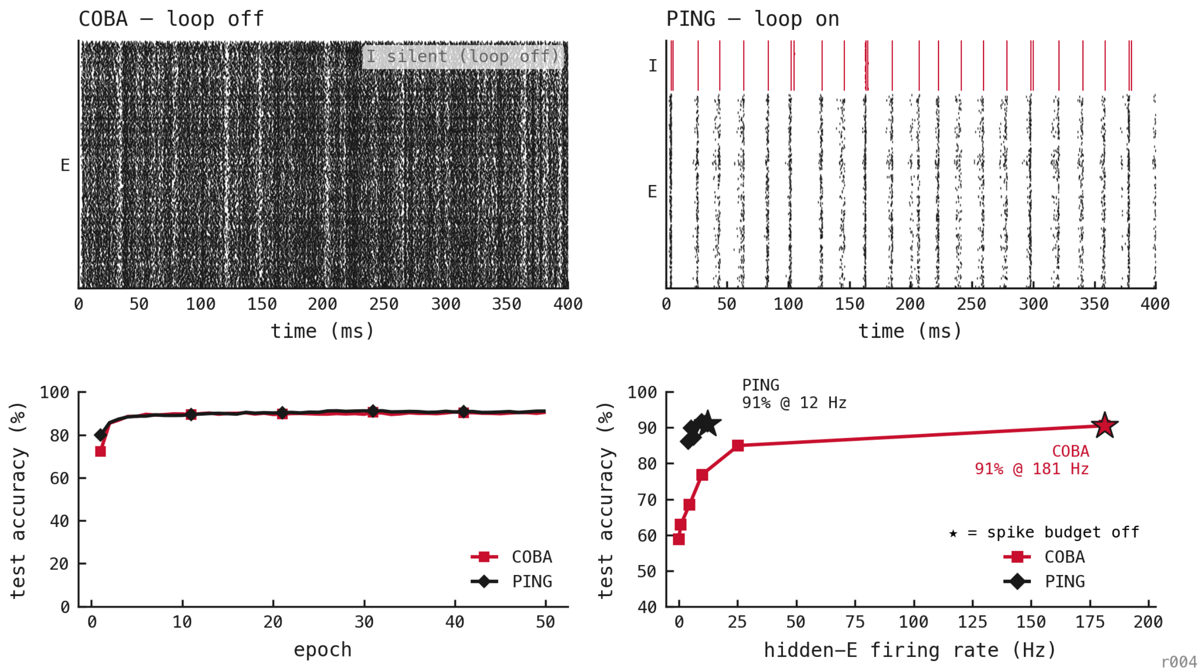


Figure 3: **Trained PING matches COBA classification accuracy while operating at an order-of-magnitude lower excitatory firing rate.** Both configurations were trained on MNIST by surrogate-gradient descent (§5.4). Top: representative single-trial rasters of the trained networks: COBA fires densely and asynchronously with the inhibitory population silent, whereas PING fires in gamma bands with synchronous inhibitory bursts (red) above excitatory spikes (black). For visualisation, each raster is an extended 400 ms replay of one digit, twice the 200 ms presentation used for training and quantitative evaluation. Bottom left: test accuracy per epoch, both configurations converging to $\approx 91\%$. Bottom right: accuracy–rate frontier traced by sweeping the per-neuron spike-budget penalty θ_u (§5.4); each marker is a trained network, plotting mean hidden-E firing rate (abscissa) against test accuracy (ordinate). PING lies above and to the left of COBA across the sweep. At the operating point with θ_u disabled (starred), PING reaches $\approx 91\%$ at ≈ 12.3 Hz, against COBA's $\approx 91\%$ at ≈ 181 Hz. Source: exp025 rate-attractor analysis in exp024.

2.4 The firing-rate reduction does not require trained loop weights

Whether the firing-rate reduction in §2.3 is acquired during training or follows from the canonical loop weights is tested by two complementary experiments. We distinguish the two senses of "architectural": (a) the reduction occurs without gradient-based tuning of the loop weights (which §2.4 establishes), as opposed to (b) the reduction emerges from generic E↔I structure without any hand-set values (which is not tested; the loop weights are held at the canonical biophysical values of [7]). The claim in this work is (a): the inductive bias is paid for at design time, not during training.

In the first, the recurrent loop is activated at inference on a trained COBA network, without retraining any weight. The network immediately produces gamma-banded rasters; the mean excitatory rate falls by approximately a factor of 15 ($\approx 133 \to\approx 9$ Hz), and test accuracy falls by ≈ 36 pp (Figure 4). The rate reduction therefore occurs without training. The accuracy loss is consistent with the absence of a learned compensation in the feedforward weights, which were optimised in the absence of the loop.

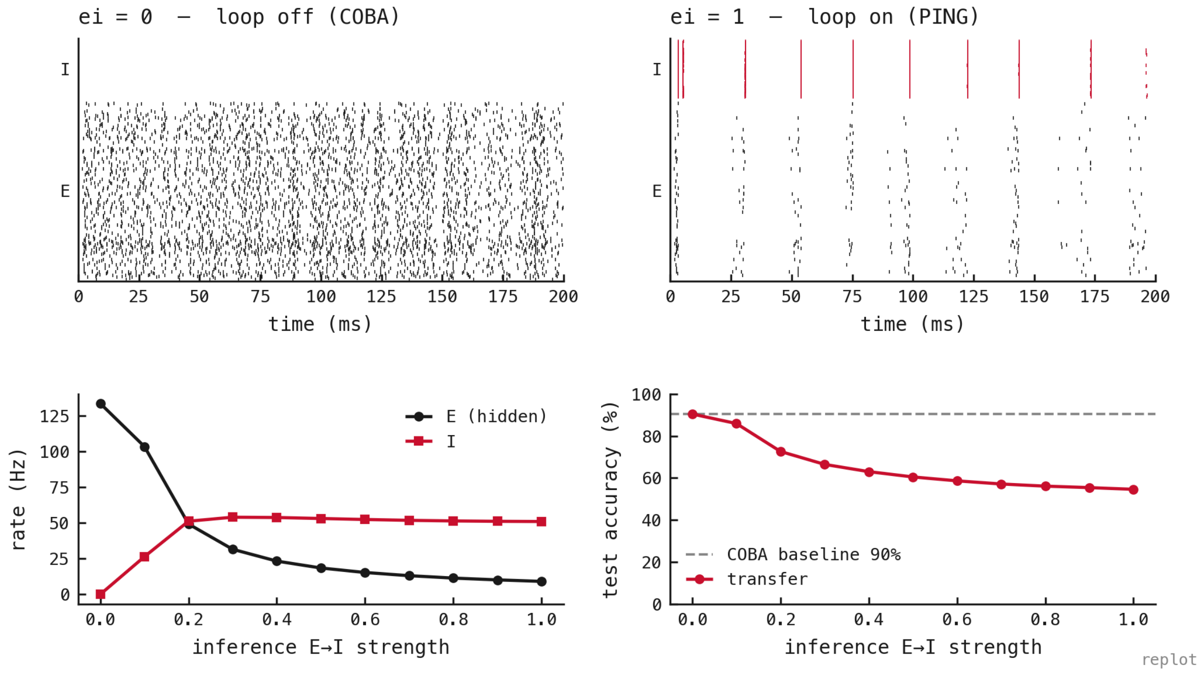


Figure 4: **Engaging the recurrent loop at inference on a trained COBA network reproduces the PING firing-rate reduction with no weight update.** A network trained in the COBA configuration is evaluated with the recurrent E→I coupling scaled from $ei = 0$ (loop off, as trained) to $ei = 1$ (canonical loop strength), without retraining. Top: single-trial rasters at $ei = 0$ (dense, asynchronous, inhibitory population silent) and $ei = 1$ (gamma bands, synchronous inhibitory bursts). Bottom left: mean excitatory (black) and inhibitory (red) firing rates versus inference coupling strength; the excitatory rate falls approximately 15-fold (from ≈ 133 to ≈ 9 Hz) as the inhibitory rate rises to ≈ 51 Hz. Bottom right: test accuracy versus coupling strength, falling from the $\approx 90\%$ COBA baseline to $\approx 55\%$ at $ei = 1$ (a ≈ 36 pp cost). The rate gating appears the instant the loop is wired in, whereas the accuracy loss reflects the absence of a feedforward compensation that only training in the presence of the loop provides. Source: exp038.

In the second, the loop weights W^{EI}, W^{IE} are released for training under the Dale's-law clamp. Both matrices represent non-negative conductance magnitudes; pathway identity fixes their reversal potentials, and the negative GABA reversal potential makes the I→E pathway inhibitory (§5.4). After each optimiser step, the trained magnitudes are projected onto the non-negative cone. In all conditions tested, the rhythmicity score collapses within a single training epoch: the first logged metric, after epoch 1, shows $R \approx 0.08$, compared with the canonical initial value $R \approx 1$ (Figure 5). The collapse is faster than the per-epoch logging interval, so no intermediate state is recorded. From every initial condition tested (canonical PING values, zero, and $0.1 \times$ canonical), the inhibitory firing rate remains near zero, R stays in 0.03–0.14, and final test accuracy is approximately 91% at ≈ 10 Hz mean E rate for the frozen-PING control (Figure 5). These numbers differ from the §2.3 frontier endpoint (≈91% at ≈12.3 Hz) because the §2.4 setup omits the spike-budget penalty θ_u and isolates the within-experiment contrast between frozen and trainable conditions rather than tracing the accuracy–rate frontier. Within this setup, gradient descent does not preserve or recover effective E→I recruitment from any tested initial condition.

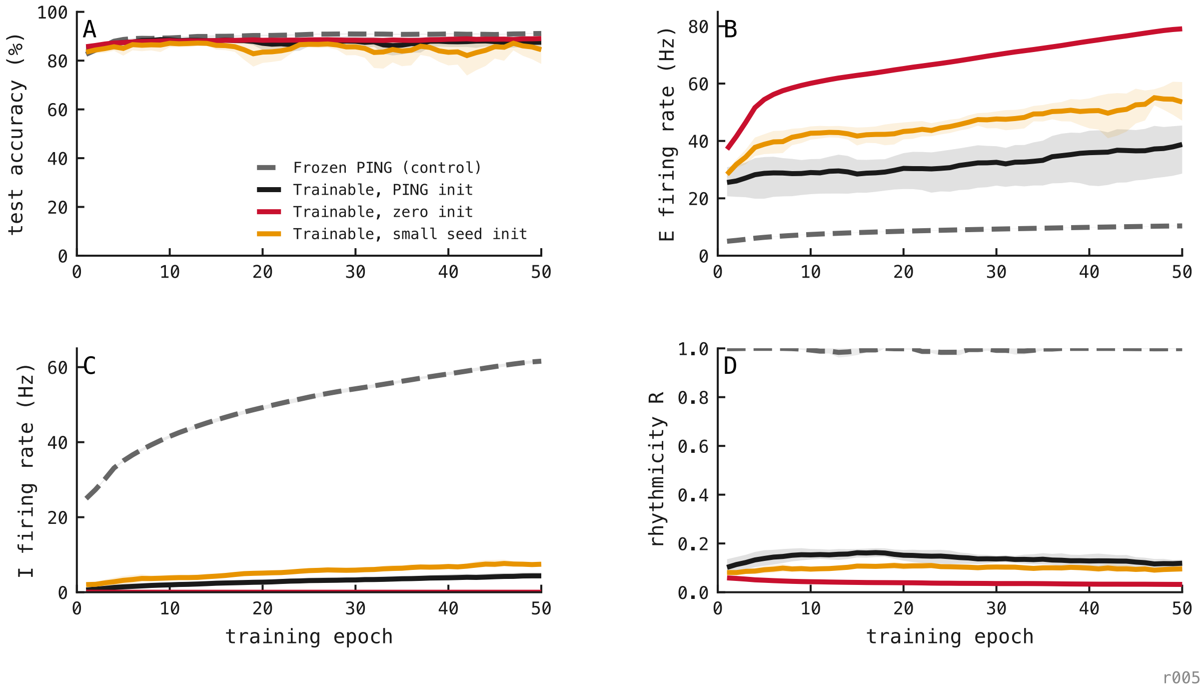


Figure 5: **Releasing the loop weights to gradient descent prunes the rhythm within a single epoch, from every initialisation.** Per-epoch training metrics with the recurrent weights W^{EI}, W^{IE} made trainable under the Dale's-law clamp, over 50 epochs on MNIST. Lines are the mean of three seeds (42–44); shading is the across-seed range. Conditions differ only in initialisation of the loop weights: canonical PING values (black), zero (red), and $0.1 \times$ canonical (amber), against a frozen-PING control (grey, dashed). **(A)** Test accuracy: all conditions overlap at ≈ 90–92%. **(B)** Mean excitatory firing rate: the trainable conditions rise to ≈ 37–76 Hz as the loop is dismantled, while the frozen control remains gated near 10 Hz. **(C)** Mean inhibitory firing rate: the trainable conditions collapse to ≈ 0 Hz within a few epochs, while the frozen control's inhibitory rate rises to ≈ 57 Hz as its readout trains. **(D)** Lobe–trough rhythmicity contrast: the frozen control holds at ≈ 1 while every trainable initialisation drains to ≈ 0.03–0.14. Source: exp049.

The second result is conditional on the gradient-damping scheme that stabilises PING training (the gradient flowing through the loop is attenuated by a factor $1/d$ on the backward pass, with $d = 1000$; §5.4); a constrained-training scheme (§4 future directions) would test whether the loop's pruning depends on the damping regime. The first experiment (inference-time activation) uses no gradients and is unaffected by this caveat, and carries most of the weight of the §2.4 conclusion.

2.5 Post-training E rate covaries with gamma frequency

The post-training excitatory firing rate covaries approximately affinely with the measured gamma frequency. Across a sweep of τ_{GABA}, which jointly changes the inhibitory decay kinetics, integrated inhibitory influence, and realised f_γ, the trained r_E is well fit by $r_E = 1.15 + 0.183 f_\gamma$ ($R^2 = 0.994$, three seeds per point; Figure 6). Mean test accuracy declines from 91.6% at $\tau_{\mathrm{GABA}} = 4.5$ ms to 86.1% at 27 ms, a 5.5 percentage-point tradeoff across the sweep. The association has a cycle-resolved counterpart. Resolving spikes by (cell, cycle) pair, 98.87% contain at most one spike across the full τ_{GABA} sweep: $P(0) \approx 79\%$, $P(1) \approx 20\%$, and the multi-spike fraction (≥ 2) is $\approx 1\%$ (Figure 7). Because τ_{GABA} changes more than frequency alone, these experiments do not identify f_γ as the sole causal variable.

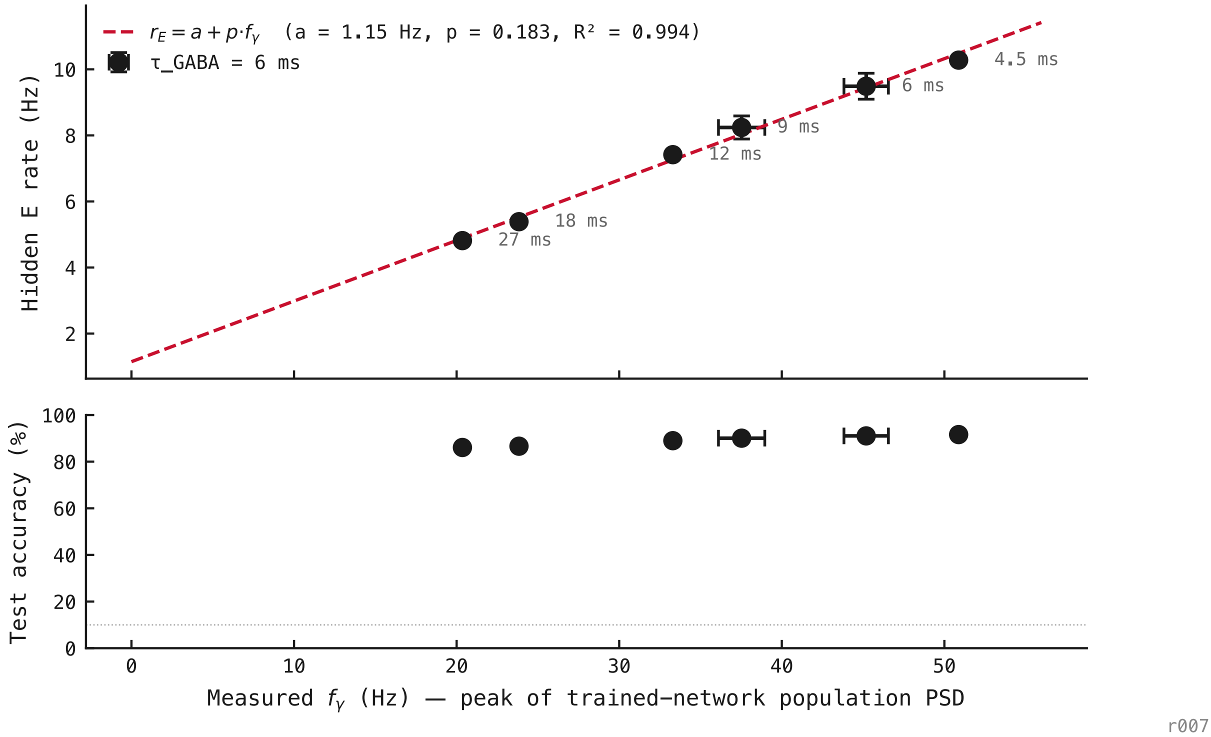


Figure 6: **Post-training excitatory rate covaries affinely with gamma frequency across a sweep of inhibitory decay kinetics.** Networks were trained from scratch at each of six values of the GABA decay constant $\tau_{\mathrm{GABA}} \in \{4.5, 6, 9, 12, 18, 27\}$ ms. Changing τ_{GABA} alters both the realised gamma frequency and the duration and integrated influence of inhibition (§5.4). Top: mean post-training excitatory firing rate against measured f_γ; markers are per-condition means over three seeds, with error bars ($\pm$ SD) on both axes. The linear fit $r_E = 1.15 + 0.183 f_\gamma$ passes through every error bar ($R^2 = 0.994$). Bottom: mean test accuracy declines from 91.6% at $\tau_{\mathrm{GABA}} = 4.5$ ms to 86.1% at 27 ms, a 5.5 percentage-point tradeoff. The sweep establishes covariance with realised frequency, not an independent causal effect of frequency alone. Source: exp041.

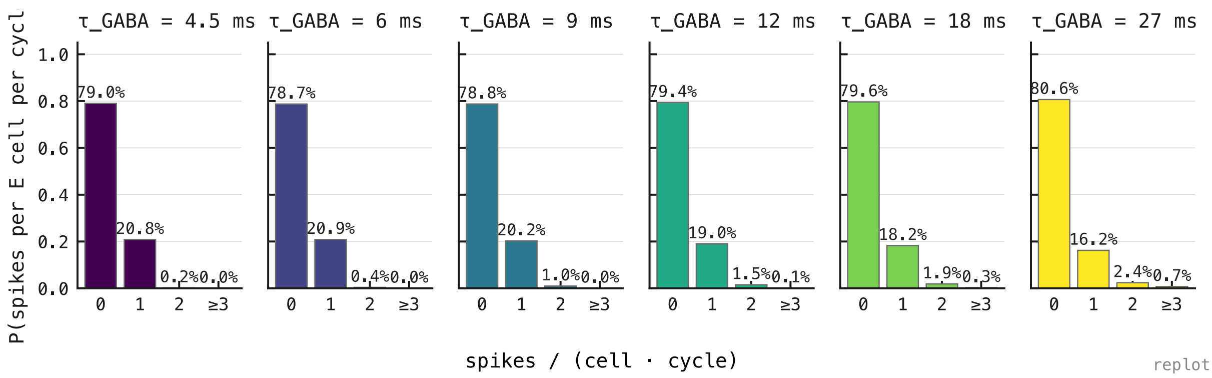


Figure 7: **The affine rate law follows from a near-binary per-cycle firing statistic: each excitatory cell contributes at most one spike per gamma cycle.** Excitatory spikes were resolved into (cell, cycle) pairs by assigning each spike to the gamma cycle inferred from peaks of the population inhibitory-burst rate (§5.5). Each panel shows the distribution of spikes per pair (0, 1, 2, or ≥ 3) at one value of τ_{GABA}. Across the sweep, $P(0) \approx 79\%$, $P(1) \approx 20\%$, and the multi-spike fraction (≥ 2) stays near $\approx 1\%$; pooled over all conditions, 98.87% of pairs contain at most one spike. Source: exp046.

2.6 Dynamics and robustness

The gating depends on the timing of inhibition, not its mean level. Perturbations of the trained PING network at inference reveal a deletion-versus-addition asymmetry. From an unperturbed baseline of approximately 91% accuracy, PING retains approximately 85% accuracy under deletion of 80% of emitted spikes (little degradation); addition of off-phase Poisson noise instead drives accuracy down to chance as the injected rate grows, because those spikes recruit the inhibitory pool at arbitrary phase and disrupt the rhythm [**31**]. COBA exhibits the opposite asymmetry (Figure 8). Two inference-time jitter perturbations of the inhibitory spike train, both holding the mean inhibitory rate fixed, produce opposite effects on the excitatory rate: per-cell jitter smears each burst into a continuous shunt and reduces the excitatory rate to zero, while cycle-coherent jitter preserves within-burst synchrony and raises the excitatory rate from ≈ 9 Hz to ≈ 36 Hz (Figure 9) [**32**].

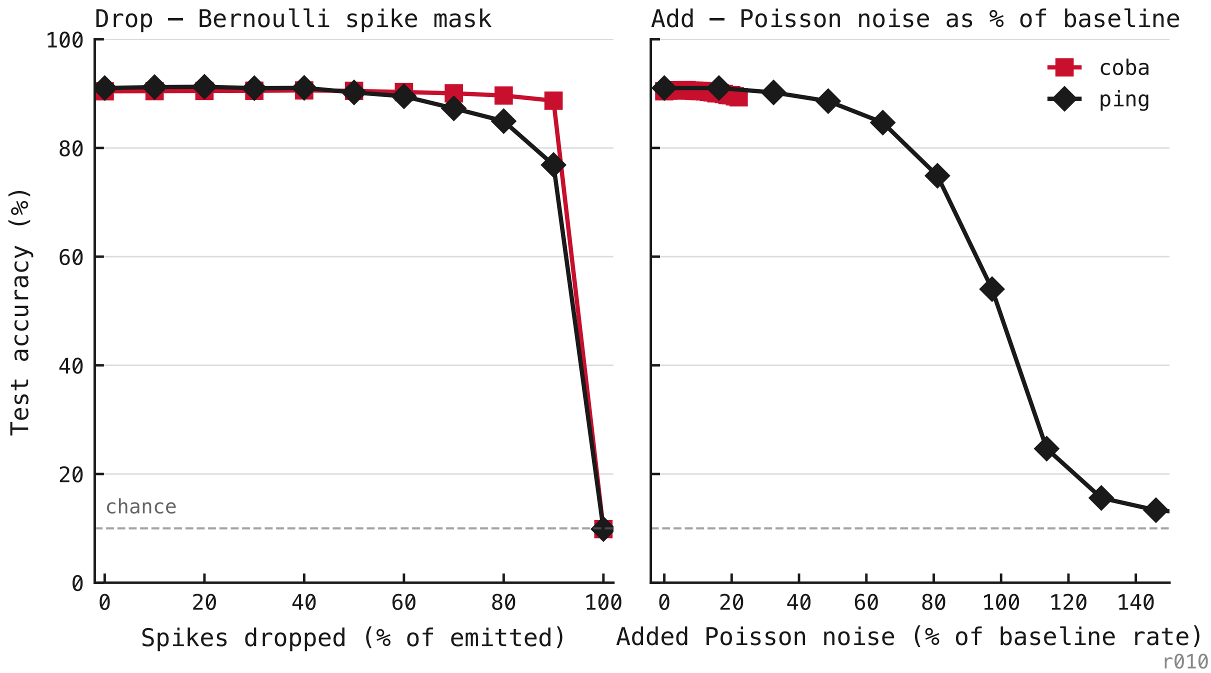


Figure 8: **The gating is robust to spike deletion but fragile to spike addition, the expected signature of a phase-based code.** A trained PING network is perturbed at inference (no retraining); accuracy is plotted against perturbation level. Left: deletion of a random fraction of emitted spikes. PING retains ≈ 85% accuracy through 80% deletion, close to its unperturbed value, degrading only as the network approaches silence. Right: addition of Poisson spikes, expressed as a fraction of each population's baseline rate. PING accuracy falls to chance as added noise grows, because off-phase spikes drive the inhibitory pool at arbitrary phase and dissolve the rhythm. COBA (red), having no rhythm to protect, shows the mirror-image asymmetry: tolerant to addition, intolerant to deletion. Source: exp037.

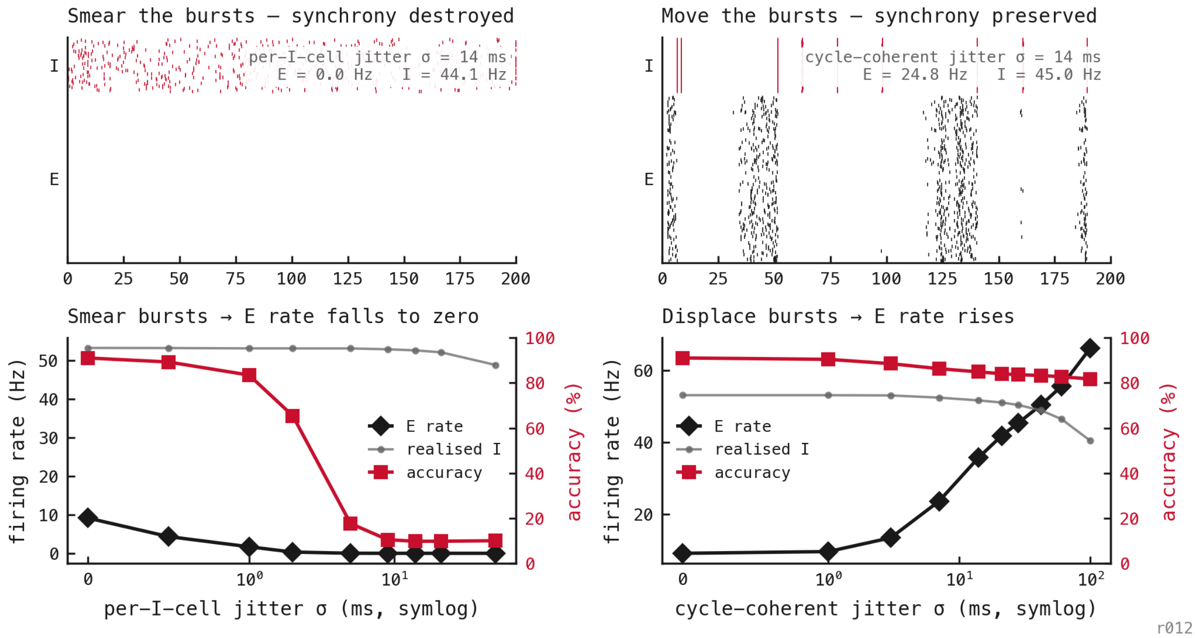


Figure 9: **Two inhibitory-jitter manipulations at the same magnitude that hold the mean inhibitory rate fixed drive the excitatory rate in opposite directions, isolating timing from level.** The trained PING inhibitory spike train is perturbed at inference while the mean per-cell inhibitory rate is held constant. Both arms use the same jitter magnitude, $\sigma = 14$ ms — only the *kind* of jitter differs. Top: single-trial rasters (E black, I red). Bottom: mean excitatory rate (black) and accuracy (red) versus jitter magnitude σ, with the realised mean inhibitory rate overlaid (grey). Left: per-cell jitter smears each burst into a continuous shunt; the excitatory rate falls to ≈ 0 Hz (inhibitory rate ≈ 53 Hz) and accuracy falls to chance. Right: cycle-coherent jitter displaces whole bursts while preserving within-burst synchrony; the excitatory rate rises from ≈ 9 to ≈ 36 Hz (inhibitory rate ≈ 52 Hz, matched to the left arm) and accuracy holds high. Both arms are read where the realised inhibitory rate is matched to within a few percent; at larger σ the cycle-coherent excitatory rate climbs further, but the finite trial window truncates the most-displaced bursts and realised inhibition falls, so the strict comparison is anchored here. Identical mean inhibition, opposite excitatory outcome: the gate is the phase structure of inhibition, not its level. Source: exp042.

The post-training firing rate is also approximately invariant under change of integration timestep: across $\Delta t \in [0.05, 1.0]$ ms (a $20\times$ range), the trained excitatory rate stays in 9.1–13.6 Hz and accuracy varies by less than 1.1 pp (Figure 10). The rate is therefore a property of the continuous dynamics, not an artefact of the discretisation.

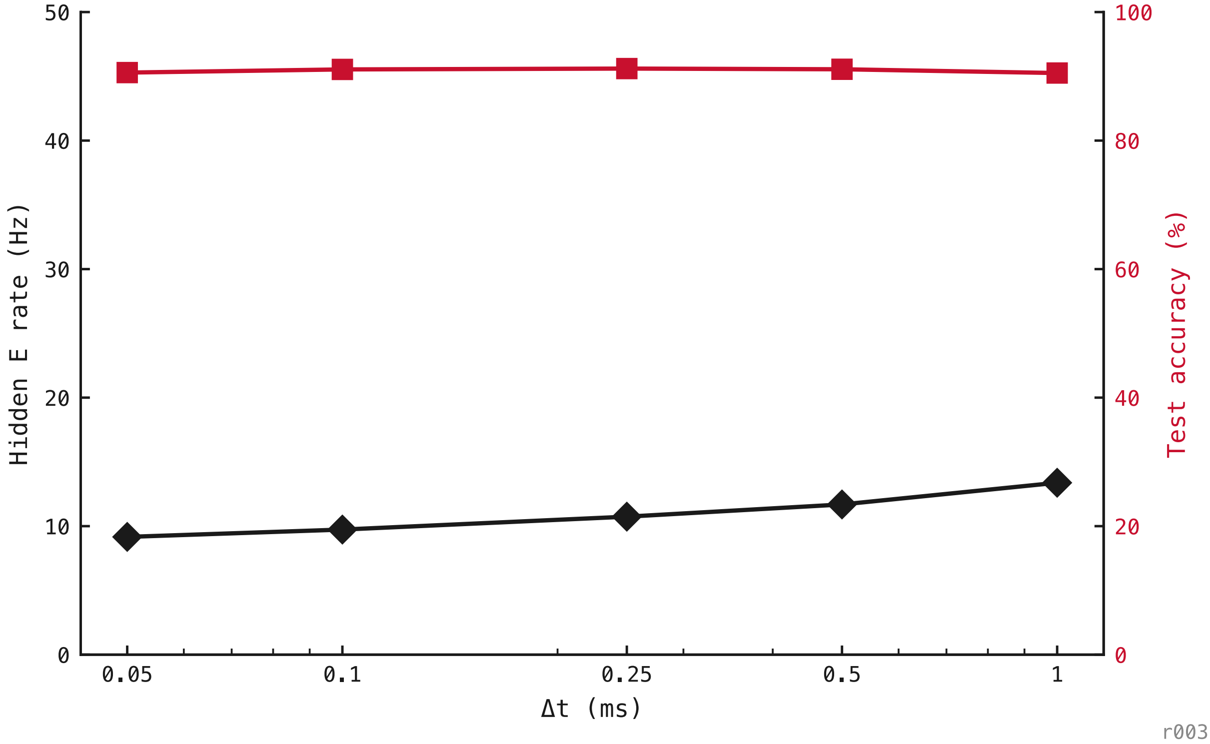


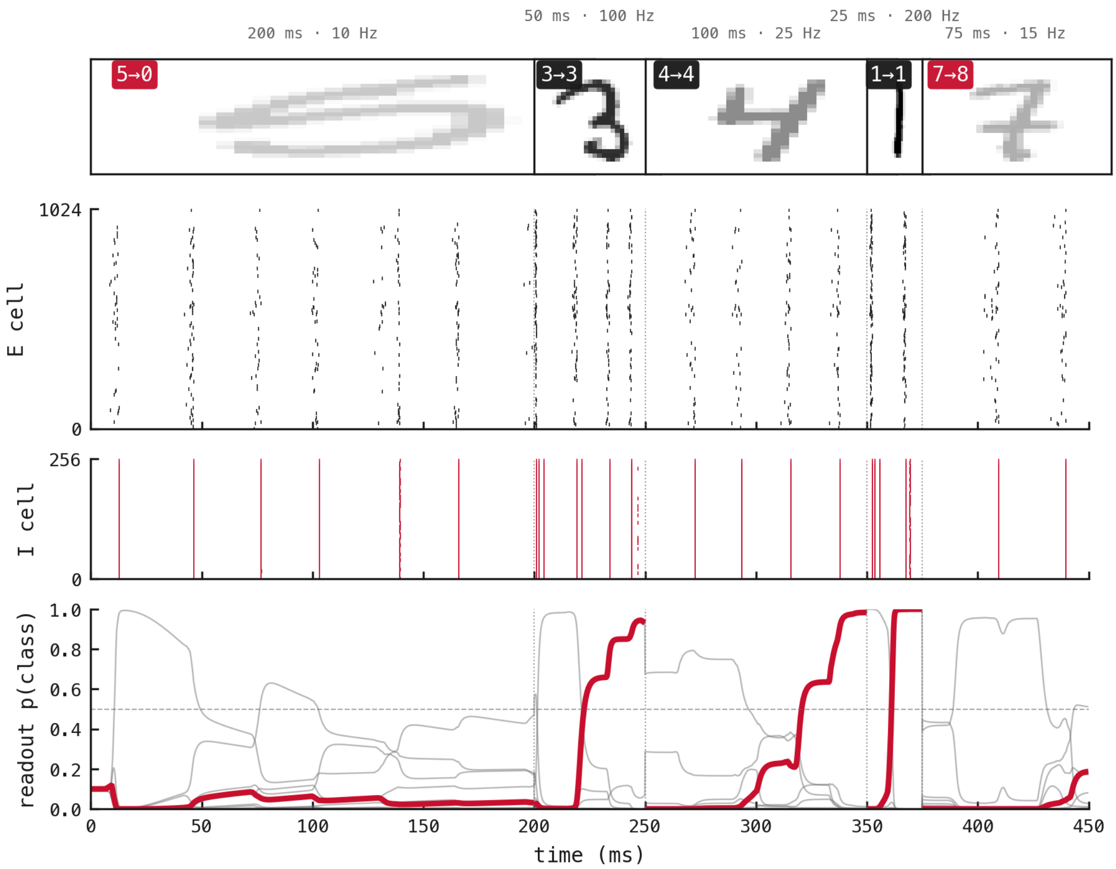
Figure 10: **The firing-rate reduction is a physical-time property, invariant to the integration timestep over a twentyfold range.** The network was trained and evaluated at matched integration timestep Δt for each value across $[0.05, 1.0]$ ms (logarithmic abscissa). Left ordinate (black diamonds): post-training mean excitatory rate, confined to a 9.1–13.6 Hz band and non-monotonic in Δt, so finer stepping does not buy a lower rate. Right ordinate (red squares): test accuracy, flat within 1.1 pp (90.4–91.4%). Because both training and inference use the same Δt at each point, the figure tests invariance of the training-plus-inference pipeline, not the generalisation of one trained network to a varied inference step. Source: exp044.

2.7 Streaming classification on continuous input

The preceding subsections evaluate the network on isolated single-digit presentations. The streaming protocol tests whether the firing-rate reduction is preserved under continuous input.

A PING network trained on single-digit MNIST classifies a continuously concatenated input stream without retraining and without a segmentation cue. The streaming protocol (§5.7) uses a time-averaged non-spiking LIF readout over a sliding window whose duration is matched exactly to each segment's presentation duration. On a representative stream of five digits, each with its own duration (25–200 ms) and input rate (10–200 Hz), 3 of 5 are classified correctly (Figure 11). Across the (τ, input-rate) grid, accuracy is approximately a function of the product $\tau \cdot$ rate. Accuracy does not exceed $\approx 80\%$ for $\tau \leq 15$ ms regardless of input rate (Figure 12A). For reference, 15 ms is approximately 0.6 times the canonical gamma period $T_\gamma \approx 27$ ms. Accuracy saturates by $\tau \approx 40$–50 ms, and the trained operating point ($\tau = 200$ ms, rate $= 25$ Hz) reaches 93% accuracy. Extending the 200 ms slice below the grid's 5 Hz minimum locates a separate encoder floor: performance remains at chance through 0.5 Hz, becomes clearly informative by 2 Hz, and reaches 79.1% at 5 Hz (Figure

12B). The failed 200 ms, 10 Hz segment in Figure 11 occurs at a population-level accuracy of 87.3%, so it is a natural weak-evidence classification error rather than evidence that the encoding rate is intrinsically nonviable. Together the panels establish requirements for sufficient integration time and sufficient encoded input evidence; because gamma frequency is not independently varied, they do not identify the gamma cycle as the cause or temporal unit of either requirement.



exp048–replot

Figure 11: **A PING network trained on isolated digits classifies a continuous, unsegmented stream whose presentation timing varies from segment to segment.** A single input stream concatenates five MNIST digits, each with its own presentation duration (25–200 ms) and Poisson input rate (10–200 Hz); the network is given no segmentation cue and is not retrained. Top: the five digit thumbnails with their per-segment (duration, rate) and predicted labels; thumbnail opacity indicates input rate. Below: hidden excitatory and inhibitory rasters, showing gamma cycles maintained throughout (sparser under weak drive, denser under strong drive) and the sliding leaky-integrator readout traces. 3 of 5 digits are classified correctly; the two errors occur in the weakest-drive segments and are interpreted against the population curve in Figure 12B. Source: exp048.

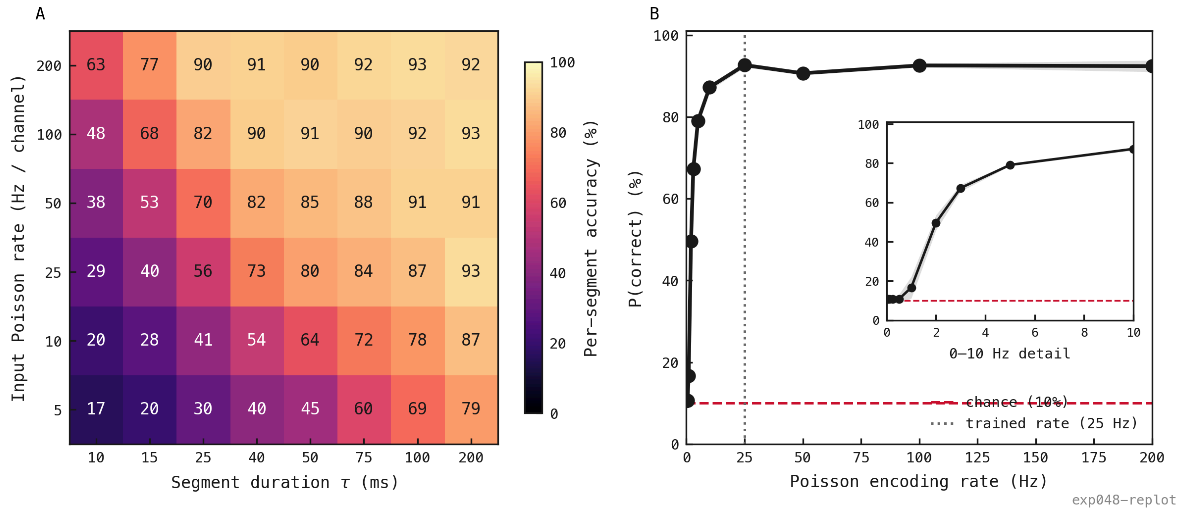


Figure 12: **Streaming accuracy has distinct integration-time and encoding-rate evidence floors. (A)** Per-segment accuracy across presentation duration τ and input rate, averaged over three seeds and 1,200 segments per grid cell. For $\tau \leq 15$ ms, accuracy does not exceed $\approx 80\%$ at any input rate; above that floor, diagonal iso-accuracy contours show an approximate dependence on $\tau \cdot$ rate. The trained operating point ($\tau = 200$ ms, 25 Hz) reaches 93%. **(B)** Probability of a correct classification versus encoding rate with both presentation duration and readout window fixed at 200 ms. All points belong to the same psychometric curve. Performance remains at chance through 0.5 Hz, is clearly informative by 2 Hz, and reaches 79.1% at 5 Hz; the dotted line marks the 25 Hz training rate. Thus the Figure 11 error at 200 ms and 10 Hz occurs in a condition with 87.3% population accuracy, above the nonviable encoder regime. For scale, $\tau = 15$ ms is approximately 0.6 times the canonical gamma period $T_\gamma \approx 27$ ms, but gamma frequency is not manipulated independently. Source: exp048.

3. Discussion

A recurrent E↔I loop, held fixed during training, generates a gamma rhythm and reduces the post-training excitatory firing rate by approximately an order of magnitude relative to the COBA baseline at matched accuracy (§2.3). The reduction does not require gradient-based learning of the loop weights: activating the loop at inference on a trained COBA network reduces the firing rate without retraining (§2.4), and gradient descent does not preserve the loop within a single epoch when its weights are released (§2.4, conditional on the §5.4 damping scheme). The inductive bias is paid for at design time, via the canonical biophysical loop weights [7], not during training. Within the mean-field reduction, the gamma onset is a supercritical Hopf bifurcation (§2.2), and the empirical data support this classification. Its continuous, reversible character fits the inference-time loop activation of §2.4, which produces a graded change in firing rate without hysteresis; a direct test would require an inference-time hysteresis sweep on the E→I gain (§4).

The mechanism of the gate is the temporal structure of inhibition, not its mean level. The jitter perturbation experiment (Figure 9) is a direct test: holding the mean per-cell inhibitory rate fixed and varying its temporal structure produces opposite effects on the excitatory rate depending on whether within-burst synchrony is preserved, consistent with prior characterisations of temporal-synchrony patterns within PING circuits [**32**]. The robustness asymmetry under spike addition versus deletion (Figure 8) follows from this dependence on phase: removal of spikes does not alter the phase structure of the population output, while addition of off-phase spikes does. The asymmetry constitutes a testable prediction for biological gamma circuits and would speak to long-standing rate-vs-timing debates [**33, 34**]. It echoes prior reports that oscillations sharpen spike-timing precision [**31**].

Across the τ_{GABA} sweep, post-training rate covaries with the realised gamma frequency, and the majority of (cell, cycle) pairs contain at most one spike (Figures 6–7). This is consistent with cycle-structured rate control, but τ_{GABA} simultaneously changes inhibitory decay, integrated conductance, and burst duty cycle; the experiment does not isolate frequency as the sole cause of the rate change. An independent manipulation of oscillation frequency at matched inhibitory influence would be needed for that attribution. The streaming experiment (Figure 12) instead separates two evidence bounds: panel A shows low accuracy at short presentation durations and an approximate dependence on $\tau \cdot$ rate, whereas panel B locates the 200 ms encoder floor below 0.5 Hz. Errors above that floor are ordinary trial-level failures under weak evidence, not evidence that the rate is categorically unusable. The numerical proximity of the short-duration floor to the canonical gamma period is descriptive, not mechanistic, because gamma frequency is not independently varied. A gamma-frequency sweep or a cycle-aligned analysis showing discontinuities at integer multiples of T_γ would be required to test whether gamma defines a temporal unit for classification.

PING is a structured alternative to the asynchronous balanced state [**36, 37**]. The architectural treatment of the loop adopted here differs from the inhibitory-plasticity literature, in which the inhibitory connectivity is plastic and learns E/I balance [**38, 39, 40, 41**]. The §2.4 result, that gradient descent does not preserve the loop from any tested initial condition (under the §5.4 damping regime), provides empirical support for the architectural treatment in this setting. We propose a functional interpretation: gamma may act as a structural constraint on excitatory firing rates without requiring learned tuning of the inhibitory connectivity.

Relative to the two closest recent trainable-SNN precedents in the bibliography, the present work differs in how the rhythm is obtained rather than in claiming better performance. [**25**] imposes the oscillation as an external input to spiking neurons; [**26**] trains an adaptive-LIF network on speech, all parameters free, and reports that oscillatory synchronisation and cross-frequency coupling *emerge* from end-to-end optimisation, correlating with task performance. The §2.4 result that gradient descent does not preserve the loop is therefore not a claim that surrogate-gradient training cannot discover rhythm in general ([**26**] shows it can) but a narrower one: it does not maintain a *fixed, biophysically-calibrated PING loop* whose weights are released under the Dale's-law clamp and the damping regime of §5.4. The two findings are complementary poles of the same question: rhythm acquired by training versus rhythm supplied by architecture. Neither [**25**] nor [**26**] uses a conductance-based E↔I loop, and neither attributes a firing-rate reduction to the rhythm via inference-time activation, frequency tuning, or jitter perturbation as §2.4–§2.6 do; their firing rates are roughly an order of magnitude higher than the rates reported here, and neither frames a per-spike economy. We do not provide a head-to-head numerical comparison: [**25**] and [**26**] evaluate on temporally structured tasks (SHD; speech perception) where the present static-MNIST protocol is not directly comparable. The contribution is mechanistic, not benchmark-driven.

The reduction reported in §2.3 should be considered net of the inhibitory contribution. PING uses a smaller, higher-rate inhibitory population, so the reduction in total spike count ($N_E\langle r_E\rangle + N_I\langle r_I\rangle$) is approximately a factor of 7 rather than a factor of 15. Excitatory glutamatergic signalling accounts for a larger share of the cortical energy budget than inhibitory transmission [**28, 29**] weighting spikes by metabolic cost recovers the order-of-magnitude figure. The argument is complicated by the substantial per-cell metabolic demands of fast-spiking interneurons, which sustain high firing rates and dense synaptic activity [**42**] for this reason we treat the uniform-spike-counting reduction (approximately 7-fold) as the more conservative claim. We do not attempt a quantitative metabolic comparison with cortex, given the architectural differences (no W_{ee} or W_{ii}, idealised synapse counts).

Several limitations apply. The evaluation uses a single dataset (MNIST), a single readout, and a fixed loop topology. MNIST is a simple, near-saturated benchmark on which many architectures reach comparable accuracy, so the classification results here should be read as evidence that the firing-rate reduction survives training to competence, not as a claim about task difficulty or about generalisation to harder problems; whether the mechanism holds on datasets with intrinsic temporal structure is left to future work (SHD, §4). The rhythmicity metric R is one of several available options. The mean-field reduction is biophysically calibrated but is a reduction of the full network. The architecture excludes W_{ee} and W_{ii}, conduction delays, and cell-type heterogeneity; the implications for cortical microcircuits with richer connectivity are open [**15**]. The streaming evaluation does not include temporally structured inputs in which classification depends on input timing. The capacity of a single gamma cycle and its scaling with assembly size are not characterised here [**43**]. Classification accuracy on MNIST is approximately 83–90%; the contribution of the present work concerns the mechanism by which the firing rate is reduced rather than the absolute accuracy. The §2.4 released-loop result depends on the gradient-damping scheme ($d = 1000$, §5.4) and the Dale clamp; whether the loop's collapse persists under weaker damping is not addressed here. The paper does not benchmark against rhythmic-SNN baselines [**25, 26**] the PING-specific attribution rests on the within-architecture experiments of §2.4–§2.6 rather than on an external rhythmic-SNN comparison.

4. Conclusion and Future Directions

A recurrent E↔I loop held fixed during training reduces the post-training excitatory firing rate by approximately an order of magnitude relative to the COBA baseline at matched accuracy, and approximately 7-fold by total spike count when the higher-rate inhibitory pool is included (§3 para 5). The reduction is invariant to the integration timestep across a $20\times$ range (Figure 10, retrained at each Δt) and is preserved under evaluation on continuously concatenated input streams (Figures 11–12).

Empirical extensions include evaluation on the SHD dataset [**44**], which tests the gating on a spiking benchmark with intrinsic temporal structure, and tasks in which classification accuracy depends on input timing, which could test whether the gamma cycle acts as a temporal unit. This is the regime in which the emergent-oscillation route of [**26**] operates; a direct comparison there would set the imposed, biophysically-calibrated PING loop studied here against a network free to learn its own rhythmic structure, on the same temporally structured task. A more comprehensive characterisation of the $W^{EI} \times W^{IE}$ plane would test the supercritical Hopf classification directly.

Theoretical extensions include multi-layer PING architectures, an independent two-dimensional sweep of the two loop-weight gains $\alpha_{EI} \times \alpha_{IE}$, multi-rhythm and theta-nested gamma models [**19, 20**], and characterisation of capacity limits for single cycles and minimal assemblies [**43**]. A constrained-training scheme, in which the loop weights are regularised toward biophysical values rather than held fixed, would connect the present result to the inhibitory-plasticity literature [**38**]. The rate law of §2.5 admits a testable biological prediction: perturbing the timing of inhibitory neurons in vivo (e.g. by optogenetic stimulation, as in [**8, 9, 10**]) should change the excitatory firing rate without altering the mean inhibitory rate, the in vivo analogue of Figure 9.

Gamma may act as a structural mechanism by which the cortical microcircuit maintains sparse excitatory firing rates at low metabolic cost. The present work provides one architecture in which this mechanism is realised explicitly.

5. Methods

5.1 Single-neuron and synapse dynamics

The model uses a conductance-based leaky integrate-and-fire (LIF) representation with two populations, excitatory (E) and inhibitory (I). The sub-threshold membrane potential of each population evolves as

$$C_m^E \frac{\mathrm{d}V^E}{\mathrm{d}t} = -g_L^E(V^E - E_L) - g_e^E(V^E - E_e) - g_i^E(V^E - E_i)$$
$$C_m^I \frac{\mathrm{d}V^I}{\mathrm{d}t} = -g_L^I(V^I - E_L) - g_e^I(V^I - E_e)$$

where C_m is the membrane capacitance, g_L the leak conductance, E_L the leak reversal potential, and E_e, E_i the excitatory and inhibitory synaptic reversal potentials. The I population has no inhibitory term because there is no I→I connection in this architecture (§5.2).

A neuron emits a spike when V crosses threshold V_{th} from below; the membrane potential is then reset to V_{reset} for a refractory period τ_{ref}:

$$s_{t+1} = \mathbf{1}[V \geq V_{\text{th}}], \quad V \leftarrow V_{\text{reset}} \text{ if } s_{t+1} = 1 \text{ or refractory.}$$

Each synaptic conductance is an exponential trace driven by presynaptic spikes; a presynaptic spike adds its full weight W as an instantaneous jump in conductance, which then decays with the relevant channel time constant (τ_{AMPA} for AMPA-like excitation, τ_{GABA} for GABA-like inhibition):

$$
\begin{aligned}
\frac{\mathrm{d}g_e^E}{\mathrm{d}t} &= -\frac{g_e^E}{\tau_{\text{AMPA}}} + W_{\text{in}} \sum_k \delta\big(t - t_k^{\text{inp}}\big) \\
\frac{\mathrm{d}g_i^E}{\mathrm{d}t} &= -\frac{g_i^E}{\tau_{\text{GABA}}} + W^{IE} \sum_k \delta\big(t - t_k^{i}\big) \\
\frac{\mathrm{d}g_e^I}{\mathrm{d}t} &= -\frac{g_e^I}{\tau_{\text{AMPA}}} + W^{EI} \sum_k \delta(t - t_k^{e})
\end{aligned}
$$

The first equation describes input-driven excitation onto E via feedforward weights W_{in}; the second is inhibition onto E from I via W^{IE}; the third is excitation onto I from E via W^{EI}. There is no equation for g_i^I (no I→I connection) and no recurrent contribution to g_e^E (no E→E connection). Canonical parameter values are listed in the parameters table; the E and I populations differ in membrane capacitance, leak conductance, membrane time constant, and refractory period.

Canonical parameter values for the spiking network are given below. Where E and I populations use different values they are listed as "E / I"; otherwise the value is shared. The $\tau_{\text{GABA}} = 9$ ms value is the canonical PING value of [**7**].

Symbol	**Description**	**Value**
C_m	Membrane capacitance	1.0 / 0.5 nF
g_L	Leak conductance	0.05 / 0.1 µS
τ_{ref}	Refractory period	3.0 / 1.5 ms
E_L	Leak reversal potential	−65 mV
V_{th}	Spike threshold	−50 mV
V_{reset}	Reset potential	−65 mV
E_e	AMPA reversal potential	0 mV
E_i	GABA reversal potential	−80 mV
τ_{AMPA}	AMPA decay time constant	2 ms
τ_{GABA}	GABA decay time constant	9 ms
Δt	Integration timestep	0.1 ms (train) / 0.25 ms (inference)
N_E	Hidden excitatory pool size	1024
N_I	Inhibitory pool size	256

5.2 Network architecture

The network has one hidden layer with N_E excitatory and N_I inhibitory units, and a non-spiking leaky-integrator readout layer with weights W_{out} over the excitatory population. The readout units follow LIF membrane dynamics with no spike, reset, or refractory period; per-class logits are computed as the membrane potentials averaged over the presentation window. Input spikes drive the excitatory population via feedforward weights W_{in}. Recurrence is restricted to the E↔I loop: E projects to I via W^{EI}, and I projects back to E via W^{IE}. There is no W_{ee} and no W_{ii}.

The restriction to the E↔I loop is intended to make the rhythm unambiguously PING: excluding W_{ii} rules out ING, and excluding W_{ee} rules out recurrent-E driven oscillation [**7,** 45]. The conductance-based (COBA) baseline used as a non-PING control is the loop-off limit of the same architecture, obtained by setting $W^{EI} = W^{IE} = 0$.

The loop weights W^{EI} and W^{IE} are held fixed (untrained) in the experiments reported in §2.3 and §2.5–§2.7. The loop is treated as a structural prior, consistent with the inhibitory-plasticity literature, in which inhibitory synapses serve an experience-dependent E/I balance role rather than carrying the feedforward computational features that the excitatory pathway acquires [**38, 39**]. The choice is also supported empirically by the result in §2.4 (Figures 4–5) that gradient descent does not preserve effective E→I recruitment when the recurrent conductances are released. A Dale's-law clamp constrains all trained conductance magnitudes to remain non-negative throughout training (§5.4).

5.3 Mean-field reduction

To locate the gamma-onset bifurcation analytically (§2.2), the spiking network is reduced to a four-dimensional rate model in the state

$$\boldsymbol{x}(t) = \left(\bar{E}, \bar{I}, \bar{g}_e^I, \bar{g}_i^E\right),$$

where $\bar{E}$ and $\bar{I}$ are the population-mean firing rates of the E and I cells (in spikes per millisecond), and $\bar{g}_e^I$, $\bar{g}_i^E$ are the population-mean cross-population synaptic conductances onto the I and E populations. The two cross-population conductances are sufficient because the architecture has no W_{ee} and no W_{ii} (§5.2); the within-population conductances vanish identically.

The reduction replaces the spike-driven conductance dynamics with rate-driven first-order filters, and replaces each cell's stochastic spike output with the population-averaged firing rate of a noise-driven LIF cell at a given mean input current. The closed 4D system is

$$\begin{aligned}
\tau_E \dot{\bar{E}} &= -\bar{E} + \Phi_E\left(I_{\text{ext}} - \bar{g}_i^E \Delta V_{\text{inh}}\right) \\
\tau_I \dot{\bar{I}} &= -\bar{I} + \Phi_I\left(\bar{g}_e^I \Delta V_{\text{exc}}\right) \\
\tau_{\text{AMPA}} \dot{\bar{g}}_e^I &= -\bar{g}_e^I + \tau_{\text{AMPA}} W^{EI} \bar{E} \\
\tau_{\text{GABA}} \dot{\bar{g}}_i^E &= -\bar{g}_i^E + \tau_{\text{GABA}} W^{IE} \bar{I}
\end{aligned}$$

where I_{ext} is an external tonic drive to E (the bifurcation control parameter, below), and the driving forces are $\Delta V_{\text{exc}} = E_e - E_L = 65$ mV and $\Delta V_{\text{inh}} = E_L - E_i = 15$ mV evaluated at rest. The membrane time constants are the passive ratios $\tau_E = C_m^E / g_L^E = 20$ ms and $\tau_I = C_m^I / g_L^I = 5$ ms, computed from the capacitances and leak conductances in the §5.1 parameters table; the synaptic time constants τ_{AMPA}, τ_{GABA} are taken directly from that table. The fan-in-normalised coupling strengths are $W^{EI} = 1.0$ µS and $W^{IE} = 2.0$ µS, inherited from the spiking network.

The population gain functions Φ_E and Φ_I are the noise-driven LIF rate functions in Ricciardi–Siegert form [**46**]. For mean input current μ delivered to a cell with leak conductance g_L, leak reversal E_L, threshold V_{th}, reset V_{reset}, membrane time constant τ_m, and refractory period τ_{ref},

$$\Phi(\mu) = \left[\tau_{\mathrm{ref}} + \tau_m\sqrt{\pi}\int_{(V_{\mathrm{reset}}-\mu_V)/\sigma_V}^{(V_{\mathrm{th}}-\mu_V)/\sigma_V} e^{u^2}(1+\mathrm{erf}\,u)\,\mathrm{d}u\right]^{-1},$$

with mean membrane potential $\mu_V = E_L + \mu/g_L$ and effective membrane-noise scale σ_V. The integral is evaluated by numerical quadrature. All parameters of Φ are taken from the §5.1 parameters table (with $\tau_m = C_m/g_L$) separately for the E and I populations. The noise scale σ_V is set to 4 mV; the predicted Hopf frequency varies by less than 1 Hz across $\sigma_V \in [3, 6]$ mV.

The silent (non-oscillating) fixed point is tracked as I_{ext} is swept from 0 to 4 nA in 10 µA steps. At each I_{ext}, the fixed point is obtained by solving the algebraic system in $(\bar{E}, \bar{I})$ (at steady state the two conductances are determined by the rates, $\bar{g}_e^I = \tau_{\mathrm{AMPA}} W^{EI}\bar{E}$ and $\bar{g}_i^E = \tau_{\mathrm{GABA}} W^{IE}\bar{I}$) using *scipy.optimize.fsolve*. The numerical Jacobian of the full 4D system at each fixed point is computed by central finite differences. The Hopf threshold $I^\star_{\mathrm{ext}}$ is the smallest I_{ext} at which the eigenvalue $\lambda^\star$ with largest real part crosses zero with non-zero imaginary part; the crossing frequency is $f^\star = |\mathrm{Im}\ \lambda^\star|\ /(2\pi)$.

The onset is classified numerically by a quasi-static amplitude sweep. I_{ext} is ramped up across $I^\star_{\mathrm{ext}}$ from a small perturbation of the silent fixed point and then ramped down; at each step the 4D system is integrated to its steady-state oscillation amplitude A. The onset is classified as supercritical when (i) the up- and down-ramp branches coincide within a peak hysteresis of 10^{-4} in rate units, and (ii) the squared amplitude scales linearly with the bifurcation distance,

$$A^2 \propto (I_{\mathrm{ext}} - I^\star_{\mathrm{ext}}),$$

with $R^2 > 0.9$. For the canonical parameter set the criterion is met with hysteresis below 10^{-5} and $R^2 = 0.999$.

The mean-field prediction is compared with the gamma frequency measured in the spiking network, extracted as in §5.5, across a sweep of $\tau_{\mathrm{GABA}} \in \{4.5, 6, 9, 12, 18, 27\}$ ms (Figure 2). Both curves decrease monotonically with τ_{GABA}; the spiking measurement is consistently higher than the rate-equation prediction across the sweep. The reduction captures the qualitative dependence on the GABA decay time, which is the use to which it is put. The treatment follows the Wilson–Cowan tradition for cortical-rhythm modelling [**18**] and the broader population-dynamics and next-generation neural-mass literature [**19, 47, 48**].

5.4 Training

The network is trained on MNIST by surrogate-gradient descent through backpropagation in time [**22, 23**]. Each digit is rate-encoded as a Poisson spike train over a 200 ms presentation window at a peak rate of 25 Hz per active pixel (Bernoulli per timestep with $p = r_{\max}\Delta t$). The loss is cross-entropy on the time-averaged membrane potentials of the non-spiking LIF readout, with a per-neuron firing-rate penalty active when a unit's mean rate exceeds a soft ceiling θ_u:

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}} + s_u \sum_{i\in E} \mathrm{ReLU}(\langle r_i\rangle - \theta_u)^2,$$

where $\langle r_i \rangle$ is the per-trial mean spike count of E neuron i, θ_u is a per-neuron rate ceiling expressed in spikes per trial, and $s_u = 10^{-3}$ is the penalty strength. Sweeping θ_u over $\{\text{off}, 5, 2, 1, 0.5, 0.2\}$ spikes per 200 ms trial (equivalent peak rates of $25, 10, 5, 2.5, 1$ Hz) generates the accuracy–rate frontier reported in §2.3.

The discrete spike nonlinearity has zero gradient almost everywhere; in the backward pass it is replaced by a fast-sigmoid surrogate of the distance to threshold $u = V - V_{\text{th}}$,

$$\frac{\partial \mathbf{1}[V \geq V_{\text{th}}]}{\partial V} \equiv \frac{s}{(1 + s\ |u|)^2},$$

with slope $s = 1$ [**23, 49**]. The forward pass evaluates the Heaviside exactly. Optimisation uses Adam [**50**] with learning rate 4×10^{-4}, batch size 256, and 50 epochs. Each baseline condition is repeated across three seeds (42, 43, 44); the spike-budget sweep uses a single seed (42).

Gradient damping for the PING configuration. Surrogate-gradient training of the PING configuration is unstable without intervention on the gradient. A single loop weight (W^{EI} or W^{IE}) contributes to the membrane-voltage update of every cell at every subsequent timestep within a trial; combined with the spike-driven impulse updates of the synaptic conductances and the non-zero surrogate gradient at the spike threshold, the gradient propagates through a tightly coupled feedback loop with millisecond-scale conductance dynamics. Over the 2,000-step backpropagation-through-time window of a single 200 ms trial at $\Delta t = 0.1$ ms, multiplicative contributions across timesteps cause gradient norms to grow by many orders of magnitude, and training does not converge. The mechanism is the same multiplicative compounding through long unrolled recurrences that characterises the exploding-gradient pathology in standard recurrent networks [**51**].

We address this by attenuating the gradient flowing through the membrane-voltage increment $\mathrm{d}V$ on the backward pass by a factor of $1/d$, implemented as a straight-through identity that scales the gradient without modifying the forward value:

$$\text{damp}_d(x) = \frac{1}{d}x + \left(1 - \frac{1}{d}\right) \text{stopgrad}(x).$$

The operator is applied to $\mathrm{d}V$ at every integration step in every layer; the forward simulation is exact (the network still solves the same ODE), and the gradient flowing through that update is attenuated by a factor of $1/d$ per step, eliminating the multiplicative compounding. All experiments reported here use $d = 1000$, applied identically to the PING and COBA training pipelines. The COBA configuration is trainable at the module default $d = 80$; the PING configuration is not.

Dale's-law clamp. The synaptic matrices store conductance magnitudes, not signed currents. After each optimiser step, W_{in}, W^{EI}, and W^{IE} are therefore projected onto the non-negative cone [**52, 53**]. Excitatory or inhibitory action is set by the pathway-specific reversal potential in the membrane current: the I→E term $g_I(E_I - V)$ is hyperpolarising for the GABA reversal potential $E_I = -80$ mV. The projection permits either recurrent conductance to grow while preventing an unphysical negative conductance. The §2.4 collapse primarily reflects weakened or absent E→I recruitment and loss of inhibitory firing, not W^{IE} crossing into a negative sign.

5.5 Measurement and analysis

Mean firing rates $\langle r_E \rangle$ and $\langle r_I \rangle$ are computed as time-averaged spike counts per neuron over the presentation window. The gamma frequency f_γ is extracted as the peak of the Welch power spectral density [**54**] of the summed population E spike train at sampling rate $f_s = 1/\Delta t = 4000$ Hz, using a single segment of length equal to the trial, mean-centred (not z-scored) signal, no detrending, and a peak search restricted to the gamma band $[5, 150]$ Hz; the peak frequency is refined by parabolic sub-bin interpolation.

The rhythmicity score R is the Michelson contrast between the first side lobe and the first trough of the autocorrelation of the binned population E spike count, computed via zero-padded FFT and normalised by the squared mean of the rate. After a 3-point smoothing kernel $[0.25, 0.5, 0.25]$ is applied to the autocorrelogram, the first trough is identified as the first local minimum starting from lag 2, and the lobe as the maximum between lag 1 and the trough; then

$$R = \frac{\text{lobe} - \text{trough}}{\text{lobe} + \text{trough}} \in [0, 1).$$

R is bounded and dimensionless. We use it in preference to spectral-peak measures because the latter become unreliable at the low firing rates encountered in some of the regimes of interest. R is qualitatively consistent with the spectral-peak and population-coherence measures used elsewhere in the gamma literature [**55, 56**].

The spikes-per-cycle distribution (Figure 7) is constructed by binning each cell's spikes into gamma cycles inferred from peaks of the population I-burst rate (Gaussian-smoothed with $\sigma = 1$ ms), detected with *scipy.signal.find_peaks* using a minimum inter-peak separation of half the expected gamma period and a height threshold of 5% of the maximum. Cycle boundaries are placed at midpoints between consecutive I-burst peaks; each E spike is assigned to its enclosing cycle.

Spike-economy claims report both the mean E rate and the total population spike count $N_E \langle r_E \rangle + N_I \langle r_I \rangle$, so the reduction is stated net of the inhibitory contribution. The metabolic argument that excitatory spikes incur larger costs than inhibitory spikes [**28, 29**] is invoked where relevant but is not modelled quantitatively.

5.6 Integration and parameters

The membrane and synaptic-conductance ODEs are integrated by an exponential-Euler scheme [**57**] with zero-order hold on the synaptic conductances over each step. With $g_{\text{tot}} = g_L + g_e + g_i$ the total instantaneous conductance, effective time constant $\tau_{\text{eff}} = C_m / g_{\text{tot}}$, and instantaneous steady state $V_\infty = (g_L E_L + g_e E_e + g_i E_i)/g_{\text{tot}}$, the closed-form update is

$$V_{t+1} = V_\infty + (V_t - V_\infty) e^{-\Delta t / \tau_{\text{eff}}}.$$

Training uses $\Delta t = 0.1$ ms; smaller timesteps are required for numerical stability of the backpropagation through the recurrent E↔I dynamics. Inference uses $\Delta t = 0.25$ ms. Firing rates and frequencies are reported in Hz. The §2.6 result (Figure 10) is obtained by retraining the network at each Δt value in the swept range, and confirms invariance of training+inference at matched Δt within the range tested (not invariance of inference at a fixed-Δt-trained network to a varied inference Δt).

Default parameters used by all experiments are listed in the parameters table. Per-experiment values of the loop weights W^{EI}, W^{IE} and the input rate, and ranges swept by experiment (e.g. τ_{GABA}, W^{IE}), are stated in the corresponding figure captions.

5.7 Datasets and evaluation

The classification task is MNIST [**58**], rate-encoded as in §5.4 (200 ms presentation, 25 Hz peak Poisson rate per active pixel). MNIST is used under its standard train/test split, and test-set accuracy is reported. For the streaming protocol used in §2.7, the trained network is presented with a continuously concatenated input stream. The output-LIF membrane potentials are averaged over a trailing window and their argmax gives the streaming class label. For every segment, the readout-window duration is matched exactly to the presentation duration, $T_{\text{readout}} = T_{\text{presentation}} = \tau$; it is not varied independently. This explicit averaging window is distinct from the output neuron's fixed membrane time constant. The duration–rate grid uses $\tau \in \{10, 15, 25, 40, 50, 75, 100, 200\}$ ms and rates in $\{5, 10, 25, 50, 100, 200\}$ Hz. Additional evaluations hold both durations fixed at 200 ms while sweeping below 5 Hz to locate the encoder's chance floor, using the same three trained seeds. Reported metrics are test accuracy, mean E and I firing rates, and f_γ.

5.8 Reproducibility

Each reported result is produced by a standalone experiment script in the project repository (linked under §6). Each experiment hardcodes its own run scale (the sample count, seed set, and any parameter sweeps) and records those settings, together with the git commit and a run identifier, in the run's provenance file; every run number quoted in this manuscript and its captions is interpolated from those files rather than typed by hand, so a figure and the text that describes it cannot drift apart. The figure-render pipeline regenerates every print-quality figure from the experiments, so a clean re-run of the repository reproduces the figures and numbers reported here.

5.9 Software and implementation

The model, training, and analysis are implemented in Python ($\geq$ 3.10). Spiking dynamics and surrogate-gradient training use PyTorch (2.11; with snnTorch 0.9 for baseline spiking primitives). Numerical analysis uses NumPy (2.2) and SciPy (1.15). All figures are produced with Matplotlib (3.10).

6. Code and data availability

Source code, per-notebook reproduction scripts, trained-weight artefacts, and the figure-render pipeline are available at https://github.com/eoinmurray/pinglab. MNIST is obtained from its standard public distributor. Library versions are listed in §5.9.

7. Declaration of generative-AI use

Claude Code (Anthropic; model Opus 4.8), an agentic large-language-model coding tool, was used extensively and interactively in the development of this work. Its use covers writing the simulation, training, analysis, and figure-rendering code in the project repository; debugging, refactoring, and code review; drafting, revising, and copy-editing the manuscript prose; and iteration on experimental design and presentation. No figure, illustration, table, or visualisation in this manuscript is produced by a generative AI model; all figures are rendered by Matplotlib from Python code in the project repository. The authors set the research questions, designed the experiments and analyses, ran the simulations, reviewed model-generated code prior to commit, and verified the reported results and the manuscript text. The authors are responsible for the content, accuracy, and claims of this work.

References

1. Buzsáki & Wang — *Mechanisms of Gamma Oscillations*. 2012. doi:10.1146/annurev-neuro-062111-150444
2. Fries — *Rhythms for Cognition: Communication through Coherence*. 2015. doi:10.1016/j.neuron.2015.09.034
3. Fries, Reynolds, Rorie & Desimone — *Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention*. 2001. doi:10.1126/science.1055465
4. Gray, König, Engel & Singer — *Oscillatory Responses in Cat Visual Cortex Exhibit Inter-Columnar Synchronization Which Reflects Global Stimulus Properties*. 1989. doi:10.1038/338334a0
5. Whittington, Traub, Kopell, Ermentrout & Buhl — *Inhibition-Based Rhythms: Experimental and Mathematical Observations on Network Dynamics*. 2000. doi:10.1016/S0167-8760(00)00173-2
6. Williams et al. — *Fast Spiking Interneurons Autonomously Generate Fast Gamma Oscillations in the Medial Entorhinal Cortex with Excitation Strength Tuning ING-PING Transitions*. 2026. doi:10.1523/ENEURO.0452-25.2026
7. Börgers — *The PING Model of Gamma Rhythms*. 2017. doi:10.1007/978-3-319-51171-9_30
8. Cardin, Carlén, Meletis, Knoblich, Zhang, Deisseroth, Tsai & Moore — *Driving Fast-Spiking Cells Induces Gamma Rhythm and Controls Sensory Responses*. 2009. doi:10.1038/nature08002
9. Sohal, Zhang, Yizhar & Deisseroth — *Parvalbumin Neurons and Gamma Rhythms Enhance Cortical Circuit Performance*. 2009. doi:10.1038/nature07991
10. Phensy et al. — *Prefrontal Gamma Oscillations Engage Dynamic Cell Type-Specific Configurations to Support Flexible Behavior*. 2026. doi:10.1016/j.neuron.2026.05.002
11. Offermanns, Pöpplau & Hanganu-Opatz — *Developmental Embedding of Parvalbumin Interneurons Drives Local and Crosshemispheric Prefrontal Gamma Synchrony*. 2026.
12. Whittington, Traub & Jefferys — *Synchronized Oscillations in Interneuron Networks Driven by Metabotropic Glutamate Receptor Activation*. 1995. doi:10.1038/373612a0
13. Wang & Buzsáki — *Gamma Oscillation by Synaptic Inhibition in a Hippocampal Interneuronal Network Model*. 1996. doi:10.1523/JNEUROSCI.16-20-06402.1996
14. Bartos, Vida & Jonas — *Synaptic Mechanisms of Synchronized Gamma Oscillations in Inhibitory Interneuron Networks*. 2007. doi:10.1038/nrn2044
15. Kopell, Börgers, Pervouchine, Malerba & Tort — *Gamma and Theta Rhythms in Biophysical Models of Hippocampal Circuits*. 2010. doi:10.1007/978-1-4419-0996-1_15
16. Viriyopase, Memmesheimer & Gielen — *Cooperation and Competition of Gamma Oscillation Mechanisms*. 2016. doi:10.1152/jn.00493.2015
17. Brunel & Wang — *What Determines the Frequency of Fast Network Oscillations with Irregular Neural Discharges? I. Synaptic Dynamics and Excitation-Inhibition Balance*. 2003. doi:10.1152/jn.01095.2002

18. Wilson & Cowan — *Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons*. 1972. doi:10.1016/S0006-3495(72)86068-5
19. Segneri, Bi, Olmi & Torcini — *Theta-Nested Gamma Oscillations in Next Generation Neural Mass Models*. 2020. doi:10.3389/fncom.2020.00047
20. Nandi, Valla & di Volo — *Bursting Gamma Oscillations in Neural Mass Models*. 2024. doi:10.3389/fncom.2024.1422159
21. Tahvili, Vinck & di Volo — *A Mean-Field Model of Neural Networks with PV and SOM Interneurons Reveals Connectivity-Based Mechanisms of Gamma Oscillations*. 2026. doi:10.1371/journal.pcbi.1014378
22. Eshraghian, Ward, Neftci, Wang, Lenz, Dwivedi, Bennamoun, Jeong & Lu — *Training Spiking Neural Networks Using Lessons From Deep Learning*. 2023. doi:10.1109/JPROC.2023.3308088
23. Neftci, Mostafa & Zenke — *Surrogate Gradient Learning in Spiking Neural Networks*. 2019. doi:10.1109/MSP.2019.2931595
24. Deckers et al. — *Advancing Spatio-Temporal Processing Through Adaptation in Spiking Neural Networks*. 2025. doi:10.1038/s41467-025-60878-z
25. Yan, Yang, Wu, Liu, Zhang, Li, Tan & Wu — *Efficient and Robust Temporal Processing with Neural Oscillations Modulated Spiking Neural Networks*. 2025. doi:10.1038/s41467-025-63771-x
26. Bittar & Garner — *Exploring Neural Oscillations During Speech Perception via Surrogate-Gradient Spiking Neural Networks*. 2024. doi:10.3389/fnins.2024.1449181
27. Barth & Poulet — *Experimental Evidence for Sparse Firing in the Neocortex*. 2012. doi:10.1016/j.tins.2012.03.008
28. Attwell & Laughlin — *An Energy Budget for Signaling in the Grey Matter of the Brain*. 2001. doi:10.1097/00004647-200110000-00001
29. Howarth, Gleeson & Attwell — *Updated Energy Budgets for Neural Computation in the Neocortex and Cerebellum*. 2012. doi:10.1038/jcbfm.2012.35
30. Ainsworth, Lee, Cunningham, Traub, Kopell & Whittington — *Rates and Rhythms: A Synergistic View of Frequency and Temporal Coding in Neuronal Networks*. 2012. doi:10.1016/j.neuron.2012.06.027
31. Schaefer, Angelo, Spors & Margrie — *Neuronal Oscillations Enhance Stimulus Discrimination by Ensuring Action Potential Precision*. 2006. doi:10.1371/journal.pbio.0040163
32. Nguyen & Rubchinsky — *Temporal Patterns of Synchrony in a Pyramidal-Interneuron Gamma (PING) Network*. 2021. doi:10.1063/5.0042451
33. Shadlen & Movshon — *Synchrony Unbound: A Critical Evaluation of the Temporal Binding Hypothesis*. 1999. doi:10.1016/S0896-6273(00)80822-3
34. London, Roth, Beeren, Häusser & Latham — *Sensitivity to Perturbations in vivo Implies High Noise and Suggests Rate Coding in Cortex*. 2010. doi:10.1038/nature09086
35. Akam & Kullmann — *Efficient "Communication through Coherence" Requires Oscillations Structured to Minimize Interference between Signals*. 2012. doi:10.1371/journal.pcbi.1002760
36. Renart, de la Rocha, Bartho, Hollender, Parga, Reyes & Harris — *The Asynchronous State in Cortical Circuits*. 2010. doi:10.1126/science.1179850

37. van Vreeswijk & Sompolinsky — *Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity*. 1996. doi:10.1126/science.274.5293.1724
38. Vogels, Sprekeler, Zenke, Clopath & Gerstner — *Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks*. 2011. doi:10.1126/science.1211095
39. Hennequin, Agnes & Vogels — *Inhibitory Plasticity: Balance, Control, and Codependence*. 2017. doi:10.1146/annurev-neuro-072116-031005
40. Wu, Miehl & Gjorgjieva — *Regulation of Circuit Organization and Function Through Inhibitory Synaptic Plasticity*. 2022. doi:10.1016/j.tins.2022.10.006
41. Páscoa dos Santos & Verschure — *Excitatory-Inhibitory Homeostasis and Bifurcation Control in the Wilson-Cowan Model of Cortical Dynamics*. 2025. doi:10.1371/journal.pcbi.1012723
42. Kann — *The Interneuron Energy Hypothesis: Implications for Brain Disease*. 2016. doi:10.1177/0271678X16638956
43. Börgers, Talei Franzesi, LeBeau, Boyden & Kopell — *Minimal Size of Cell Assemblies Coordinated by Gamma Oscillations*. 2012. doi:10.1371/journal.pcbi.1002362
44. Cramer, Stradmann, Schemmel & Zenke — *The Heidelberg Spiking Data Sets for the Systematic Evaluation of Spiking Neural Networks*. 2022. doi:10.1109/TNNLS.2020.3044364
45. Tiesinga & Sejnowski — *Cortical Enlightenment: Are Attentional Gamma Oscillations Driven by ING or PING?*. 2009. doi:10.1016/j.neuron.2009.09.009
46. Brunel — *Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons*. 2000. doi:10.1023/A:1008925309027
47. Gerstner — *Population Dynamics of Spiking Neurons: Fast Transients, Asynchronous States, and Locking*. 2000. doi:10.1162/089976600300015899
48. Montbrió, Pazó & Roxin — *Macroscopic Description for Networks of Spiking Neurons*. 2015. doi:10.1103/PhysRevX.5.021028
49. Zenke & Ganguli — *SuperSpike: Supervised Learning in Multilayer Spiking Neural Networks*. 2018. doi:10.1162/neco_a_01086
50. Kingma & Ba — *Adam: A Method for Stochastic Optimization*. 2015.
51. Pascanu, Mikolov & Bengio — *On the Difficulty of Training Recurrent Neural Networks*. 2013.
52. Cornford, Kalajdzievski, Leite, Lamarquette, Kullmann & Richards — *Learning to Live with Dale's Principle: ANNs with Separate Excitatory and Inhibitory Units*. 2021.
53. Zhu et al. — *Task Success in Trained Spiking Neural Network Models Coincides with Emergence of Cross-Stimulus-Modulated Inhibition*. 2026. doi:10.1007/s00422-025-01030-4
54. Welch — *The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms*. 1967. doi:10.1109/TAU.1967.1161901
55. Atallah & Scanziani — *Instantaneous Modulation of Gamma Oscillation Frequency by Balancing Excitation with Inhibition*. 2009. doi:10.1016/j.neuron.2009.04.027
56. Xing, Shen, Burns, Yeh, Shapley & Li — *Stochastic Generation of Gamma-Band Activity in Primary Visual Cortex of Awake and Anesthetized Monkeys*. 2012. doi:10.1523/JNEUROSCI.5644-11.2012

57. Rotter & Diesmann — *Exact Digital Simulation of Time-Invariant Linear Systems with Applications to Neuronal Modeling*. 1999. doi:10.1007/s004220050570
58. LeCun, Bottou, Bengio & Haffner — *Gradient-Based Learning Applied to Document Recognition*. 1998. doi:10.1109/5.726791