Temporal and spatial evidence limits of trained PING

exp048 · 8 June 2026

Abstract

A frozen pyramidal-interneuron gamma (PING) network, whose trained weights remain fixed during evaluation, is tested under complementary temporal and spatial reductions of Modified National Institute of Standards and Technology (MNIST) digit evidence. It classifies continuously streamed digits without retraining, but a duration × input-rate sweep reveals a failure floor below 15 ms. At fixed 200 ms presentation and readout windows, performance remains at 10-class chance through 0.5 Hz and becomes clearly informative by 2 Hz. Separately, foreground pixels are permanently removed from binarized images and presented to both PING and a width-matched artificial neural network (ANN). PING is competitive at intermediate deletion but reaches chance by retention q = 0.02. Together, the curves delimit temporal, event-rate, and spatial evidence regimes for a future variable-rate training experiment.

Methods

Streaming duration and encoding rate

The trained baseline from the canonical PING experiment contains 1024 excitatory (E) and 256 inhibitory (I) cells. It was trained on one MNIST digit per 200 ms trial using random seeds 42, 43, 44. Each seed identifies an independent training run with a reproducible random initialization and data order. Everything here is **inference-only** at the trained timestep $\Delta t = 0.1$ ms; the weights are never updated. The two-dimensional sweep averages over all 3 seeds; the single-stream examples use seed 42.

A stream of digits, each shown for τ ms, is classified in one forward pass. The *only* change from training is a sliding readout window:

1. **Encode** each digit as a Poisson spike train over τ ms across 784 input channels, one per pixel. Each channel generates independent random spikes. The stated encoding rate is the expected number of spikes per second for a full-intensity pixel; lower pixel intensities reduce that rate proportionally.
2. **Concatenate** the per-digit trains into one input stream. At segment index k, the Poisson rate switches instantaneously at the boundary

$$t_k = k\tau. \quad (1)$$

   Here t_k is the boundary time of segment k, k is the segment index, and τ is the duration of each segment.

3. **Run once** through the trained network at the trained Δt, without retraining.

4. **Integrate evidence** in a non-spiking output leaky integrator, one unit per class. A leaky integrator accumulates incoming spikes while gradually discounting older input:

$$v_{\text{out}}(t) = \beta_{\text{out}} v_{\text{out}}(t-1) + \frac{1-\beta_{\text{out}}}{\Delta t} \boldsymbol{s}^{E}(t-1) W_{\text{out}}. \quad (2)$$

   Here t is the discrete timestep; v_{out} (t) is the vector of output-unit states; β_{out} is their leak factor, the fraction of the previous output state retained for one timestep; Δt is the simulation timestep; s^{E} (t−1) is the E-cell spike vector at the preceding timestep; and W_{out} is the trained E-to-output weight matrix.

5. **Read a sliding window.** Average v_{out} over the trailing τ-window and apply a softmax:

$$\text{logits}(t) = \frac{\Delta t}{\tau} \sum_{u=t-w+1}^{t} v_{\text{out}}(u). \quad (3)$$

$$p(\text{class}, t) = \text{softmax}(\text{logits}(t)). \quad (4)$$

   Here logits(t) is the class-evidence vector; u indexes timesteps in the window; τ is the presentation duration; w is its number of timesteps; and p(class,t) is the softmax-normalized class-probability vector. Softmax converts the logits into non-negative class probabilities that sum to one. The readout-window duration is matched exactly to the current digit's presentation duration:

$$T_{\text{readout}} = T_{\text{presentation}} = \tau. \quad (5)$$

   Here $T_{readout}$ is the duration over which output evidence is averaged, $T_{presentation}$ is the time for which the digit is shown, and τ is that common duration.

   The corresponding window length is

$$w = \frac{\tau}{\Delta t}. \quad (6)$$

   Every digit is therefore read over exactly its own presentation duration; readout duration is not varied independently. At training the average ran over the *whole* trial; the trailing matched-duration window is the single change.

6. **Predict per segment** at the end of the digit's τ-window according to

$$\hat{c}(t) = \arg\max_{c} p(\text{class} = c, t). \quad (7)$$

   Here c indexes the 10 digit classes, and arg max selects the class with the largest probability.

The output leak is

$$\beta_{\text{out}} = \exp(-\Delta t / \tau_{\text{out}}). \quad (8)$$

Here τ_{out} is the output-unit time constant, which controls how quickly accumulated output evidence decays, and exp denotes the exponential function. The probability trace p(class,t) is the network's online class confidence.

The grid uses presentation durations 10 ms, 15 ms, 25 ms, 40 ms, 50 ms, 75 ms, 100 ms, 200 ms and input rates 5 Hz, 10 Hz, 25 Hz, 50 Hz, 100 Hz, 200 Hz per channel. This gives 48 cells with 1200 classified segments per cell.

To resolve the encoding-rate floor below the grid, additional evaluations use rates 0.01 Hz, 0.025 Hz, 0.05 Hz, 0.1 Hz, 0.25 Hz, 0.5 Hz, 1 Hz, 2 Hz, 3 Hz while holding both presentation and readout at 200 ms. Each cell contains 10 streams of 10 digits for every trained seed. The 5 Hz, 10 Hz, 25 Hz, 50 Hz, 100 Hz, 200 Hz points use the same fixed-duration protocol and come from the corresponding grid row.

Foreground-retention calibration

The spatial protocol uses the same MNIST split and has two parts:

1. Train 3 seeds of a width-matched artificial neural network (ANN) with 784 inputs, one rectified-linear hidden layer of 1024 units, and 10 outputs. The hidden width matches the PING E population, not its recurrent E/I architecture. A rectified-linear unit outputs zero for a negative input and otherwise passes the input unchanged. Training uses 15 epochs, or complete passes through the training set, batches of 256 images per weight update, and learning rate 0.001, the step size of each update.
2. Binarize each held-out image, meaning an image excluded from training, at intensity 0: pixels above the threshold become unit-valued foreground and all others become zero-valued background. Retain every foreground pixel independently with probability q. Retention $q = 1$ leaves the foreground intact and $q = 0$ removes it. The ANN calibration uses 10 independent mask realizations per image. The matched comparison uses 100 fixed held-out examples and identical masks for every ANN and PING seed; PING encodes them at 25 Hz for 200 ms.

Results

Streaming classification and temporal evidence

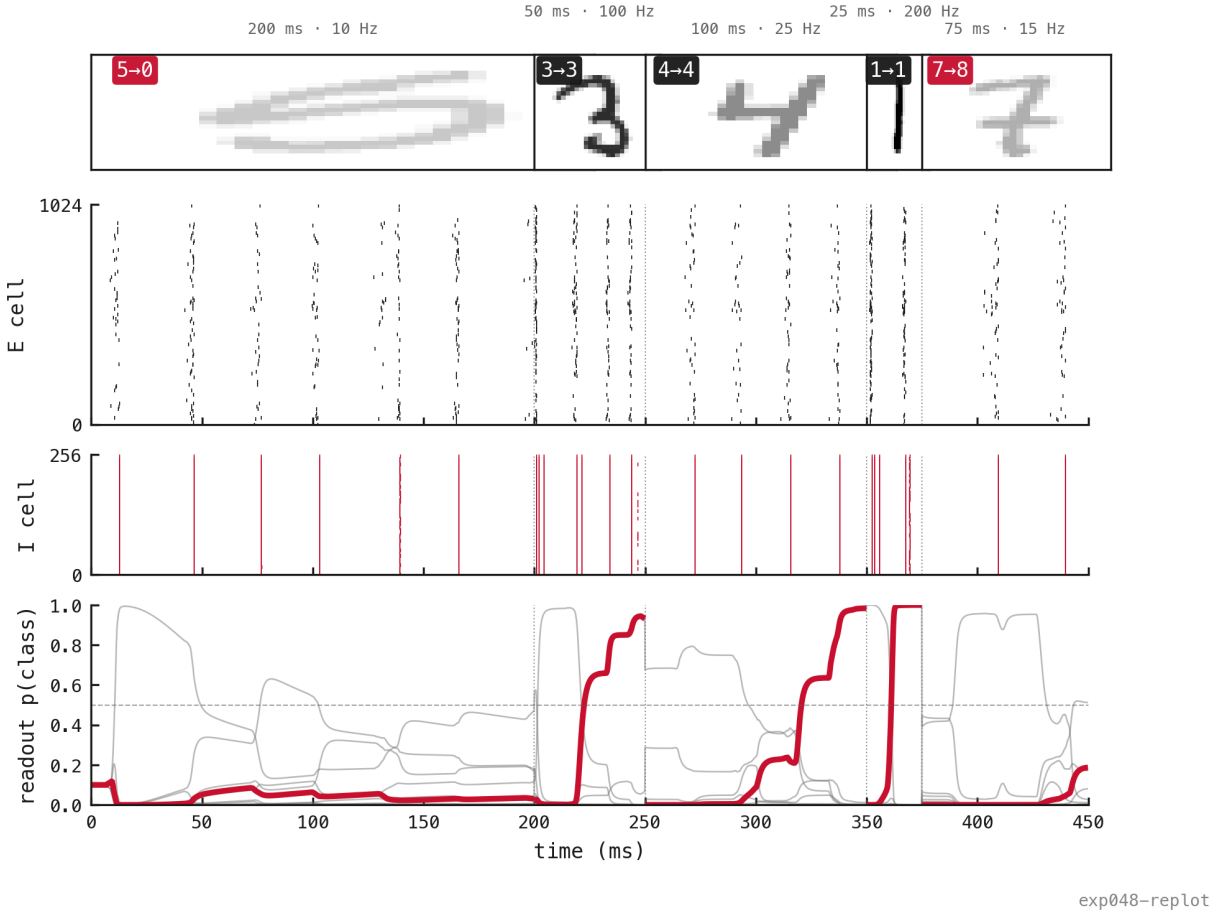

Figure 1: Classification when presentation duration and encoding rate vary between segments. The segment conditions are 200 ms at 10 Hz; 50 ms at 100 Hz; 100 ms at 25 Hz; 25 ms at 200 Hz; 75 ms at 15 Hz. Thumbnail opacity increases with encoding rate. The middle panels plot E- and I-cell spike rasters against time (ms); the lower panel plots class probability against time (ms), with the true class emphasized in red. The label-to-prediction pairs are 5→0, 3→3, 4→4, 1→1, 7→8, giving 3 of 5 correct segments.

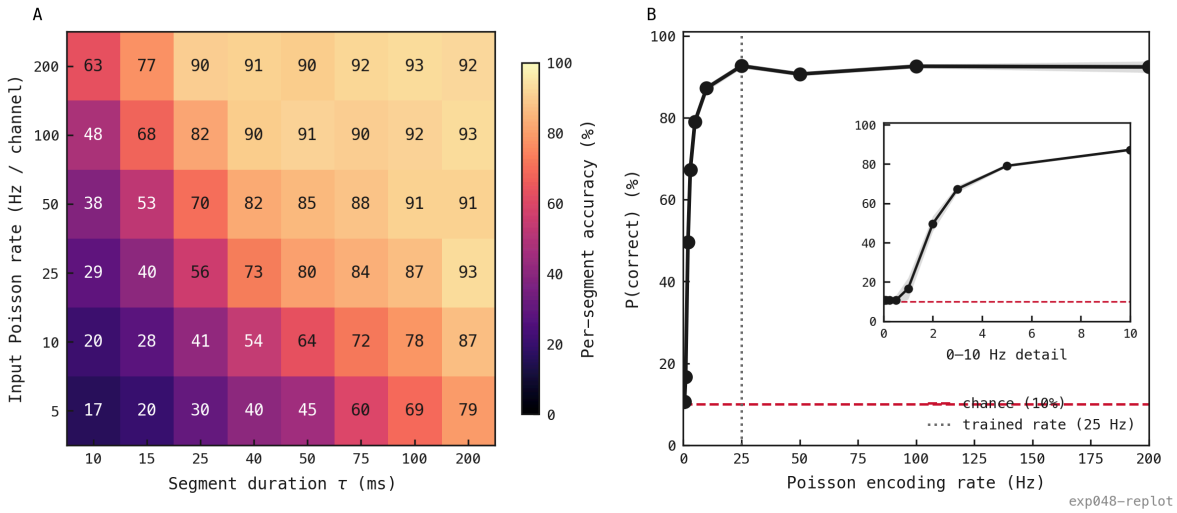

Figure 2: Temporal and encoding-rate limits of the frozen PING classifier. **(A)** Per-segment accuracy (%) is shown for presentation duration (ms, horizontal) and Poisson encoding rate (Hz per channel, vertical), using 1200 segments per cell. **(B)** Probability of a correct classification (%) is plotted against encoding rate (Hz) with presentation and readout fixed at 200 ms. The inset enlarges the linear 0.01–10 Hz interval without changing the axis scale. The dashed line marks 10-class chance and the dotted line the 25 Hz training rate. Accuracy stays at its empty-input floor, the accuracy obtained when almost no input spikes arrive, through 0.5 Hz, becomes informative by 2 Hz, and reaches 79.1% at 5 Hz.

The fixed-duration rate curve distinguishes a nonviable encoder regime from ordinary classification errors under weak evidence. In the variable-condition stream, the first failed segment received 10 Hz for 200 ms, yet that condition reaches 87.3% across the population. Its error is therefore natural trial-level variation, not evidence that 10 Hz is intrinsically too low. The other failed segment, presented at 15 Hz for 75 ms, is likewise above the empty-input rate floor, although its shorter window supplies less total evidence. Rates below 0.5 Hz are not useful operating points; 2 Hz is the lowest clearly informative tested rate and 5 Hz is a practical lower bound for future sweeps.

Spatial evidence calibration

The architecture-matched ANN remains above chance until foreground retention falls to q = 0.005, fewer than one visible foreground pixel per image on average.

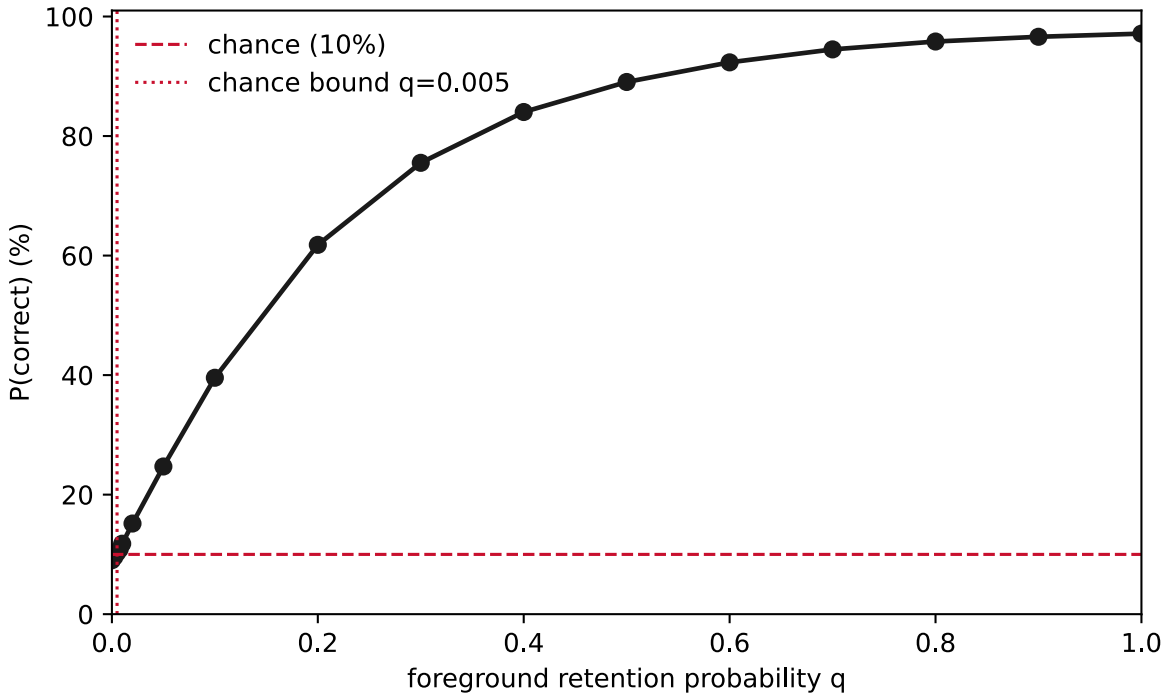

Figure 3: Held-out ANN accuracy as foreground evidence is removed. Probability of a correct classification (%) is plotted against foreground retention q, the independent probability that a foreground pixel remains visible. Points are means across 3 ANN seeds and the band is one standard error, the estimated uncertainty of that mean across seeds. The dashed line marks 10-class chance; the dotted line marks the measured chance-region bound at q = 0.005, the highest tested retention whose 95% confidence interval across seeds still contains chance accuracy.

Under identical masks, neither classifier is uniformly better. With 29.5 visible pixels on average (q = 0.2), PING reaches 68.7% against the ANN's 60.7%. They coincide near 45% at q = 0.1. PING still leads at q = 0.05, then reaches chance at q = 0.02 while the ANN remains above it.

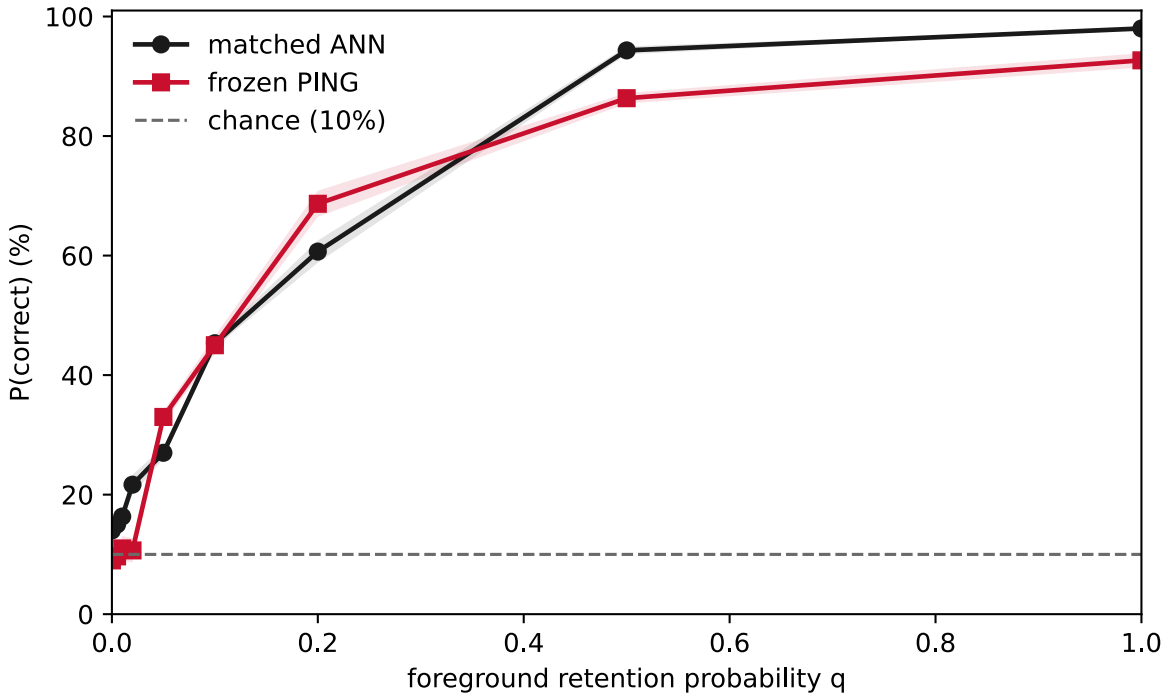

Figure 4: Width-matched ANN and frozen PING accuracy under identical spatial deletion. Probability of a correct classification (%) is plotted against foreground retention q. Black circles denote ANN and red squares PING; bands show one standard error across trained seeds. Each point uses 100 fixed held-out examples and identical independently sampled foreground masks. PING runs at 25 Hz for 200 ms. Neither classifier dominates across the full retention range.

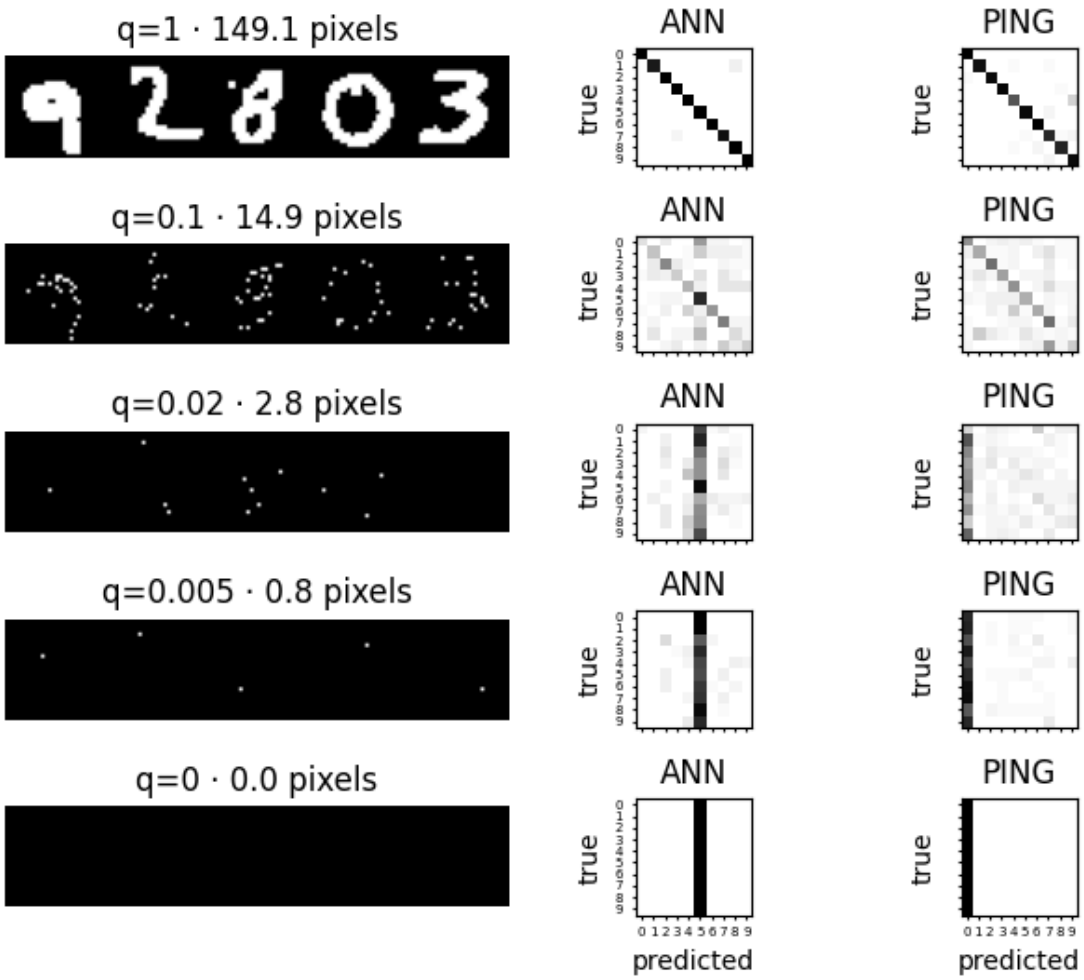

Figure 5: Stimulus and error structure across the matched masking curve. Rows progress from intact input through intermediate masking to the blank control. Left panels show five binarized examples and their mean visible foreground-pixel count. The ANN and PING panels show confusion matrices with true digit on the vertical axis and predicted digit on the horizontal axis; each row is normalized to show the distribution of predictions for one true class. The lowest-retention rows reveal collapse toward model-specific default classes rather than structured digit confusions.

Binarization maps every nonzero antialiased MNIST pixel, including intermediate-intensity pixels introduced to smooth digit edges, to full intensity, so spatial retention is not a second measurement of grayscale contrast. Nevertheless, expected input-event count gives a useful first-order bridge. For an otherwise identical binary image encoded at the masking experiment's 25 Hz ceiling, retaining fraction q gives the same expected event count as retaining the full foreground and using

$$r_{\text{equiv}} = q \cdot 25 \ \text{ Hz}. \quad (9)$$

Here r_{equiv} is the full-foreground Poisson rate with the same expected event count, meaning the expected total number of input spikes across the image and presentation; q is foreground retention; and 25 Hz is the masking experiment's reference encoding rate.

Thus q = 0.1 maps to r_{equiv} = 2.5 Hz, within the 2–3 Hz transition of the fixed-duration rate curve. At that retention, PING reaches 45%, compared with 49.7% at 2 Hz in the grayscale rate sweep. This numerical alignment supports including a rate near 2.5 Hz in variable-rate training, but it is not an equivalence of corruptions: spatial masking permanently removes locations, whereas lowering Poisson rate preserves all locations in expectation and changes temporal sampling noise.