

Supplemental Material for

Polychronous Wave Computing: Timing-Native Address Selection in Spiking Networks

Natalia G Berloff, DAMTP, University of Cambridge
n.g.berloff@damtp.cam.ac.uk

This supplement specifies (i) notation, (ii) the competitor statistic K_{eff} , and (iii) full numerical protocols for Figs. 2 – 7.

Contents

S1 LUT-SNN formalism and notation	2
S2 Correlated template libraries and an effective number of competitors	2
S2.1 Runner-up statistic and definition of K_{eff}	3
S2.2 Alternative estimate from template overlaps	4
S2.3 Numerical protocol for Fig. 2 (linear discrimination margin)	5
S3 Extended nonlinear simulations: multi-port WTA and 1D slices	6
S3.1 Template library, inputs, and disorder	7
S3.2 Nonlinear competitive digitizer	7
S3.3 Monte Carlo accuracy estimation	8
S4 Four-mode driven–dissipative model for the temporal-order comparator	8
S4.1 Modes and equations of motion	8
S4.2 Four-mode junction Hamiltonian	9
S4.3 Pulse drives	9
S4.4 Initialization and integration	9
S4.5 Integrated-energy readout	9
S4.6 Three-spike permutation decoding and cycle repair (Fig. 5)	10
S5 Calibration and hardware-in-the-loop optimization	10
S5.1 Static disorder model (Fig. 6)	11
S5.2 Comparator readout	12
S5.3 Calibration objective	12
S5.4 SPSA update rule	12
S5.5 Training and evaluation protocol (Fig. 6)	12
S5.6 Measurement noise and limits	13

S6 End-to-end MoE routing benchmark: numerical details	13
S6.1 Setup: hard top-1 routing with injected errors	13
S6.2 Synthetic data and labels	13
S6.3 Experts and training	14
S6.4 Routing-error injection	14
S6.5 Device-to-benchmark mapping	14
S6.6 Evaluation scope and cascading	14
S7 End-to-end scaling snapshot (interface-limited baseline vs. wave-domain selection).	15

S1 LUT-SNN formalism and notation

This section fixes notation for the lookup-table spiking neural network (LUT-SNN) layer used in the main text and isolates the role of the *interferometric polychrony primitive*. At the abstraction level used throughout the Supplement, one device instance implements a fast K -way address selection (winner-take-all across K output ports) from a spatiotemporal spike pattern, while the *value readout* associated with the selected address can be performed either optically (all-wave) or electronically (hybrid). Table 1 fixes the symbols used below. One evaluation cycle is specified by N event times

$$x \equiv (t_1, \dots, t_N), \quad 0 \leq t_j \leq t_{\max}, \quad (\text{S1})$$

where the index j labels a fixed input channel (or synapse), and t_j is the spike time on that channel; j does not denote temporal order. We represent times in a rotating frame by unit phasors

$$u_j \equiv e^{-i\Omega t_j}, \quad (\text{S2})$$

and collect them into $u = (u_1, \dots, u_N)^\top$. Conceptual aspects of the phase-delay mapping (shift invariance, wrap-free window, and coherence constraints) are given in the main text (Secs. II–IV); here we use only one immediate consequence: a global time shift $t_j \mapsto t_j + t_0$ multiplies all phasors by a common factor $u_j \mapsto e^{-i\Omega t_0} u_j$ and therefore leaves any intensity-based decision rule invariant.

S2 Correlated template libraries and an effective number of competitors

Template libraries are rarely i.i.d.: shared structure, clustering, or common-mode disorder induces correlations among *mismatched* channels. To summarize how strongly incorrect channels compete with the correct one, we introduce an *operational* effective competitor count K_{eff} , defined so that a simple independent-competitor reference model reproduces the observed *runner-up* (nearest-competitor) statistics. With this definition, correlated libraries can yield $K_{\text{eff}} < K$, while strongly *crowded* near-tie landscapes can yield $K_{\text{eff}} > K$.

The independent-template “random-phase” model is used only as a *reference* limit. In the low-jitter, weak-disorder regime relevant to Fig. 2, unmatched contribution phases $\varphi_{jk} \equiv \theta_{jk} - \phi_j$ (with $\phi_j = \Omega t_j$) are not i.i.d. uniform: they exhibit strong clustering. Accordingly, Rayleigh/random-phase amplitude nulls are not used to *fit* the runner-up distribution in this regime. Instead, we report (i) the operational crowding metric K_{eff} derived from runner-up near-tie statistics (Sec. S2.1) and (ii) a spectral diversity proxy $K_{\text{eff}}^{\text{spec}}$ computed from template overlaps (Sec. S2.2).

Symbol	Meaning (as used in the Supplement)
N	Number of input spikes/events within one decision window
t_j	Time of event j within the window ($j = 1, \dots, N$)
t_{\max}	Maximum decision-window duration
Ω	Rotating-frame angular frequency; wrap time $T_{\text{wrap}} = 2\pi/\Omega$
ϕ_j	Phase encoding, $\phi_j = \Omega t_j$
u_j	Unit phasor, $u_j = e^{-i\phi_j} = e^{-i\Omega t_j}$
\mathbf{u}	Input phasor vector $(u_1, \dots, u_N)^\top$
K	Number of templates/classes (ports) in one address-selection module
K_{eff}	Effective competitor count: the number of <i>independent</i> competitors
k	Class/port index, $k = 1, \dots, K$
\mathbf{w}_k	Complex template (weight) vector for class k
J_{jk}	Coupling from input j to template/port k ; $J_{jk} \equiv (\mathbf{w}_k)_j$ (often $J_{jk} = J e^{i\theta_{jk}}$)
Ψ_k	Complex interferometric score for class k (linear in \mathbf{u})
$I_k^{(\text{lin})}$	Linear intensity score, $I_k^{(\text{lin})} = \Psi_k ^2$
I_k	Context-dependent intensity score: $ \Psi_k ^2$ (linear scorer) or $ \psi_k(t_{\text{read}}) ^2$ (WTA readout)
\widehat{k}_{lin}	Linear decision rule, $\widehat{k}_{\text{lin}} = \arg \max_k I_k^{(\text{lin})}$
k^*	Ground-truth class (correct address)
$\Delta_{\text{lin}}^{\log}$	Log-intensity margin, $\log(I_{k^*}^{(\text{lin})} / \max_{k \neq k^*} I_k^{(\text{lin})})$
σ_t	Timing jitter standard deviation
σ_θ	Static phase disorder standard deviation
$\psi_k(t)$	WTA competitive state variable for channel k
P_{WTA}	WTA selection accuracy under specified noise/disorder
P_{lin}	Linear baseline accuracy under specified readout noise

Table 1: Compact notation map used throughout the Supplement.

S2.1 Runner-up statistic and definition of K_{eff}

For notational simplicity we omit the LUT layer index and write the K complex scores as $\Psi_k(x)$. For an input x whose correct address is k^* , define the runner-up amplitude

$$A_{\text{ru}}(x) \equiv \max_{k \neq k^*} A_k(x), \quad A_k(x) \equiv |\Psi_k(x)|. \quad (\text{S3})$$

We define K_{eff} as the value (not necessarily an integer) that makes an independent-competitor reference model match the *near-tie* statistics of A_{ru} observed in simulation/experiment. The random-phase analysis below is included only to fix notation and provide a baseline runner-up CDF; deviations from it are summarized by K_{eff} .

Reference runner-up CDF (random-phase null). We first define the uniform-coupling reference model $|J_{jk}| = J$ used in Sec. S2.3. Writing $J_{jk}u_j = J e^{i(\theta_{jk} - \phi_j)}$ with $\phi_j = \Omega t_j$, the contribution from input j to template k has total phase $\varphi_{jk} \equiv \theta_{jk} - \phi_j$. Under the null model for an *unmatched* template, the phases $\{\varphi_{jk}\}_{j=1}^N$ are taken to be independent and approximately uniform on $[0, 2\pi)$; we suppress the template index and write φ_j below.

For an unmatched template, the phasor sum $\Psi_k = \sum_{j=1}^N J e^{i\varphi_j}$ has approximately circular-

Gaussian real and imaginary parts, so $A_k = |\Psi_k|$ is approximately Rayleigh with

$$\Pr(A_k \leq a) \approx 1 - \exp\left(-\frac{a^2}{NJ^2}\right). \quad (\text{S4})$$

Assuming $K - 1$ independent unmatched competitors, the runner-up amplitude CDF is

$$\Pr(A_{\text{ru}} \leq a) \approx \left[1 - \exp\left(-\frac{a^2}{NJ^2}\right)\right]^{K-1}. \quad (\text{S5})$$

For non-uniform couplings one replaces NJ^2 by the corresponding variance scale $\sum_j |J_{jk}|^2$ (or its typical value across k).

Operational estimate (near-tie crowding from runner-up ratios). Given labeled inputs $\{(x^{(n)}, k^{*(n)})\}$, define for each sample the matched (winner) and strongest-incorrect (runner-up) *linear* intensities

$$I_{\text{win}}^{(n)} \equiv I_{k^{*(n)}}^{(\text{lin})}(x^{(n)}), \quad I_{\text{ru}}^{(n)} \equiv \max_{k \neq k^{*(n)}} I_k^{(\text{lin})}(x^{(n)}), \quad (\text{S6})$$

and form the runner-up ratio

$$r^{(n)} \equiv \frac{I_{\text{ru}}^{(n)}}{I_{\text{win}}^{(n)}} \in [0, 1]. \quad (\text{S7})$$

Let $F_{\text{emp}}(r)$ denote the empirical CDF of $\{r^{(n)}\}$.

To compress near-tie crowding into a single scalar, we fit F_{emp} to a one-parameter order-statistics reference family: if r_{max} is the maximum of M i.i.d. $\text{Uniform}([0, 1])$ variables, then $\Pr(r_{\text{max}} \leq x) = x^M$. We therefore use the model CDF $F_\kappa(x) = x^{\kappa-1}$ (with $\kappa \equiv M + 1$) and define

$$K_{\text{eff}} \equiv \arg \min_{\kappa \in [1, \kappa_{\text{max}}]} \sup_{x \in [0, 1]} |F_{\text{emp}}(x) - x^{\kappa-1}|, \quad \kappa_{\text{max}} \gg K, \quad (\text{S8})$$

i.e. the best-fit κ under a Kolmogorov–Smirnov discrepancy. Interpreted this way, K_{eff} is an operational *near-tie crowding* metric (not a literal count): it summarizes correlations, non-uniform coupling norms, and common-mode disorder through their effect on runner-up ratios.

S2.2 Alternative estimate from template overlaps

The runner-up procedure yields an *operational* crowding parameter K_{eff} from data. Here we give a complementary, data-free diagnostic that depends only on the stored template phases and measures the *global diversity* of the library.

Normalized template vectors. For each class k , define a unit-norm phase template $v_k \in \mathbb{C}^N$ by

$$(v_k)_j \equiv \frac{1}{\sqrt{N}} e^{i\theta_{jk}}, \quad \|v_k\|_2 = 1, \quad (\text{S9})$$

(or $(v_k)_j \propto |J_{jk}| e^{i\theta_{jk}}$ with normalization when magnitudes are non-uniform).

Template-overlap (Gram) matrix. Define the squared overlaps

$$G_{kk'} \equiv |v_k^\dagger v_{k'}|^2 = \frac{1}{N^2} \left| \sum_{j=1}^N e^{i(\theta_{jk} - \theta_{jk'})} \right|^2, \quad G_{kk} = 1, \quad G_{kk'} \in [0, 1], \quad (\text{S10})$$

so large off-diagonal entries indicate redundant (highly correlated) templates.

Table 2: **Template diversity in the main figures.** The spectral effective count $K_{\text{eff}}^{\text{spec}}$ is computed from the participation ratio of the template Gram matrix G [Eqs. (S11)], and satisfies $1 \leq K_{\text{eff}}^{\text{spec}} \leq K$. For Fig. 2 we report $K_{\text{eff}}^{\text{spec}}$ separately for the *diverse* (unsorted i.i.d.) and *crowded* (sorted order-statistic) template ensembles used in panels (a) and (b), respectively.

Figure / panel	K	N	$K_{\text{eff}}^{\text{spec}}$	$K_{\text{eff}}^{\text{spec}}/K$
Fig. 2(a) diverse (unsorted i.i.d.)	8	32	7.25 ± 0.11	0.91
Fig. 2(b) crowded (sorted)	8	32	1.13 ± 0.03	0.14
Fig. 3 (WTA envelope)	6	10	5.51 ± 0.09	0.92

Spectral effective count. We define a *spectral effective template count* as the participation ratio of G ,

$$K_{\text{eff}}^{\text{spec}} \equiv \frac{(\text{tr } G)^2}{\text{tr}(G^2)} = \frac{K^2}{\sum_{k,k'} G_{kk'}^2}. \quad (\text{S11})$$

This lies in $[1, K]$: $K_{\text{eff}}^{\text{spec}} = K$ for orthogonal libraries ($G = I$) and $K_{\text{eff}}^{\text{spec}} \rightarrow 1$ when all templates are effectively identical.

How to interpret $K_{\text{eff}}^{\text{spec}}$ versus K_{eff} . The two effective counts quantify different notions of “competition.” $K_{\text{eff}}^{\text{spec}}$ measures *global library diversity* (how many distinct phase patterns are stored), whereas K_{eff} measures *local decision crowding* (how close the strongest competitor typically lies to the winner for the scoring/noise conditions of interest). Because K_{eff} is defined from runner-up statistics, it is not constrained by K and can exceed it in strongly crowded regimes. Accordingly, we report both numbers: $K_{\text{eff}}^{\text{spec}}$ as a redundancy diagnostic, and K_{eff} as the quantity most directly tied to nearest-competitor errors.

In summary, $K_{\text{eff}}^{\text{spec}} \approx K$ indicates a diverse (weakly correlated) library, whereas $K_{\text{eff}}^{\text{spec}} \approx 1$ indicates a globally low-diversity (redundant) library. Figure 2 uses these two controlled ensembles to separate two effects: (i) coherence loss with timing jitter, which produces smooth margin decay in both cases, and (ii) *crowding*, where a small zero-jitter margin makes device-level phase disorder shift the decision boundary. Accordingly, the “operating-window penalty” reported for Fig. 2(b) is a property of the crowded ensemble, not the diverse one.

S2.3 Numerical protocol for Fig. 2 (linear discrimination margin)

This subsection specifies the simulation protocol for Fig. 2 (linear log-margin versus timing jitter), including static per-coupling phase disorder.

Linear scoring model. We simulate a single interferometric scoring stage with $K = 8$ templates and $N = 32$ input spikes. For an input $x = (t_1, \dots, t_N)$, the complex score at port k is

$$\Psi_k(x) = \sum_{j=1}^N \exp[i(\theta_{jk} - \Omega t_j)], \quad (\text{S12})$$

with linear intensities $I_k^{(\text{lin})} = |\Psi_k|^2$ and log-margin

$$\Delta_{\text{lin}}^{\log} = \log \left(\frac{I_{k^*}^{(\text{lin})}}{\max_{k \neq k^*} I_k^{(\text{lin})}} \right). \quad (\text{S13})$$

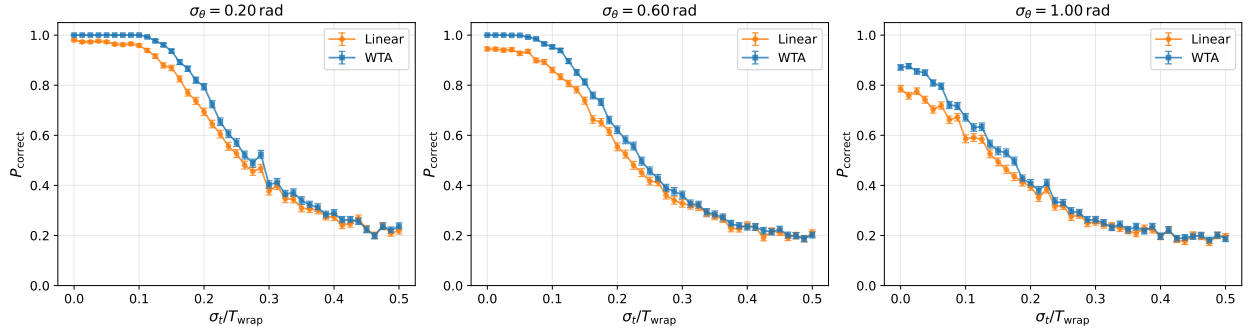


Figure 1: One-dimensional cuts through the operating envelope. Classification accuracy P_{correct} versus timing jitter σ_t/T_{wrap} for fixed phase disorder $\sigma_\theta \in \{0.20, 0.60, 1.00\}$ rad, comparing the linear baseline decision \hat{k}^{lin} [Eq. (S21)] and the nonlinear WTA decision \hat{k}^{WTA} [Eq. (S23)] under the Monte Carlo protocol of Sec. S3.3. Error bars are binomial standard errors computed from $n_{\text{dev}}n_{\text{trial}} = 1000$ samples per point.

Template libraries and timing window. We set $\Omega/(2\pi) = 10$ GHz so that $T_{\text{wrap}} = 2\pi/\Omega = 100$ ps and choose a wrap-free window $\Delta t_{\text{max}} = 0.6 T_{\text{wrap}} = 60$ ps. Compiled phases are $\theta_{jk} = \Omega t_{jk}^{\text{templ}}$. To generate the two Fig. 2 regimes we use: (i) *diverse* libraries with i.i.d. spike times $t_{jk}^{\text{templ}} \sim \text{Unif}(0, \Delta t_{\text{max}})$ (unsorted), and (ii) *crowded* libraries obtained by sorting the N times within each pattern (order statistics), which produces strongly correlated templates. (No minimum inter-spike separation is enforced.)

Inputs and timing jitter. For each trial, the ground-truth class k^* is sampled uniformly from $\{1, \dots, K\}$ and the query is generated by

$$t_j = \text{clip}_{[0, \Delta t_{\text{max}}]}(t_{jk^*}^{\text{templ}} + \epsilon_j), \quad \epsilon_j \sim \mathcal{N}(0, \sigma_t^2), \quad (\text{S14})$$

with $\sigma_t \in \{0, 2, 4, 6, 8, 10, 12, 15, 18, 22, 26, 30, 35, 40, 45, 50\}$ ps.

Static phase disorder. Per-coupling phase offsets are applied as

$$\theta_{jk} \rightarrow \theta_{jk} + \delta\theta_{jk}, \quad \delta\theta_{jk} \sim \mathcal{N}(0, \sigma_\theta^2), \quad (\text{S15})$$

with $\sigma_\theta = 0.2$ rad. For each device realization (template bank), the full set $\{\delta\theta_{jk}\}$ is fixed across all trials and jitter levels.

Monte Carlo. For each σ_t , we run 200 trials per class (1600 trials per jitter level) and repeat over 5 independent template banks (seeds 42–46), for a total of 128,000 trials per regime (diverse vs crowded). Points in Fig. 2 report the mean and SEM across the 5 seeds.

S3 Extended nonlinear simulations: multi-port WTA and 1D slices

This section specifies the numerical protocol for the multi-port WTA results in the main text (Sec. IV) and Fig. 3. We simulate a single address-selection module and therefore drop the layer/table index, writing Ψ_k for the k -th score. We describe (i) template/input generation with disorder, (ii) the reduced driven–dissipative digitizer, and (iii) Monte Carlo accuracy estimation.

S3.1 Template library, inputs, and disorder

Template library. For each address $k \in \{1, \dots, K\}$ and input $j \in \{1, \dots, N\}$, draw template times

$$t_{jk}^{\text{templ}} \sim \text{Unif}(0, t_{\text{max}}), \quad (\text{S16})$$

and hold the full library $\{t_{jk}^{\text{templ}}\}$ fixed while sweeping $(\sigma_t, \sigma_\theta)$.

Compiled phases and device disorder. With uniform magnitudes $|J_{jk}| \equiv 1$, compile phases as $\theta_{jk} = \Omega t_{jk}^{\text{templ}}$. A device realization $d \in \{1, \dots, n_{\text{dev}}\}$ draws static offsets

$$\theta_{jk}^{(d)} = \theta_{jk} + \delta\theta_{jk}^{(d)}, \quad \delta\theta_{jk}^{(d)} \sim \mathcal{N}(0, \sigma_\theta^2), \quad (\text{S17})$$

independently over (j, k) and fixed across all trials for that device.

Inputs with timing jitter. Each trial samples $k^* \sim \text{Unif}\{1, \dots, K\}$ and generates

$$t_j = \text{clip}_{[0, t_{\text{max}}]}(t_{jk^*}^{\text{templ}} + \epsilon_j^{\text{jit}}), \quad \epsilon_j^{\text{jit}} \sim \mathcal{N}(0, \sigma_t^2), \quad (\text{S18})$$

independently over j , and we write $x \equiv (t_1, \dots, t_N)$.

Interferometric scores (seeds). Given (x, d) , compute

$$\Psi_k(x; d) = \sum_{j=1}^N \exp\left[i\left(\theta_{jk}^{(d)} - \Omega t_j\right)\right]. \quad (\text{S19})$$

Linear baseline with readout noise. For the linear baseline we apply multiplicative amplitude noise,

$$\tilde{\Psi}_k = \Psi_k(x; d) (1 + \epsilon_k), \quad \epsilon_k \sim \mathcal{N}(0, \sigma_{\text{readout}}^2), \quad \sigma_{\text{readout}} = 0.25, \quad (\text{S20})$$

with independent real ϵ_k across k , and decide

$$\hat{k}_{\text{lin}} = \arg \max_{k \in [K]} |\tilde{\Psi}_k|. \quad (\text{S21})$$

Ties are broken by choosing the smallest index.

S3.2 Nonlinear competitive digitizer

To digitize the seed vector $\{\Psi_k\}$ we integrate the reduced competition model

$$\frac{d\psi_k}{dt} = \left[G - \gamma - \eta |\psi_k|^2 - \chi \sum_{\ell \neq k} |\psi_\ell|^2 \right] \psi_k + \alpha \Psi_k, \quad k = 1, \dots, K, \quad (\text{S22})$$

with parameters

$$\gamma = 1, \quad G = 1.2, \quad \eta = 1, \quad \chi = 1, \quad \alpha = 0.4.$$

Unless noted otherwise we use

$$\Omega = 2\pi, \quad T_{\text{wrap}} = 1, \quad t_{\text{max}} = 0.9, \quad K = 6.$$

For Fig. 3A we take $N = 40$ (visualization density), and for the operating-envelope maps in Fig. 3B–F we take $N = 10$.

We integrate Eq. (S22) with RK4 (step $dt = 10^{-2}$) to $t_{\text{final}} = 20$ and read out at $t_{\text{read}} = 0.3$:

$$\widehat{k}_{\text{WTA}} = \arg \max_{k \in [K]} |\psi_k(t_{\text{read}})|^2. \quad (\text{S23})$$

Ties are broken by choosing the smallest index. For diagnostics only, we additionally check whether the instantaneous winner remains unchanged over the final fraction $f = 0.1$ of the integration interval; runs failing this check are counted as incorrect.

S3.3 Monte Carlo accuracy estimation

For each $(\sigma_t, \sigma_\theta)$ we use $n_{\text{dev}} = 4$ device realizations and $n_{\text{trial}} = 250$ trials per device, for $n_{\text{dev}}n_{\text{trial}} = 1000$ samples per grid point. Accuracies are computed as

$$P_{\text{lin}} = \Pr(\widehat{k}_{\text{lin}} = k^*), \quad P_{\text{WTA}} = \Pr(\widehat{k}_{\text{WTA}} = k^*).$$

Figure S1 shows that WTA extends the usable jitter window at fixed σ_θ relative to the noisy linear baseline, while the two methods coincide in the high-accuracy and high-noise limits.

S4 Four-mode driven–dissipative model for the temporal-order comparator

This section specifies the four-mode junction model and numerical protocol used for the temporal-order comparator (Fig. 4) and its reuse in the three-spike cascade of Fig. 5. Performance metrics are defined in Supplement Sec. S3; here we document the dynamics, parameters, and readout.

S4.1 Modes and equations of motion

We model two driven input modes A, B and two coupled readout modes L, R by

$$\boldsymbol{\psi}(t) = (\psi_A, \psi_B, \psi_L, \psi_R)^\top \in \mathbb{C}^4.$$

Dynamics follow the driven complex-envelope equation

$$\frac{d\boldsymbol{\psi}}{dt} = -iH\boldsymbol{\psi} - iU(|\boldsymbol{\psi}|^2 \odot \boldsymbol{\psi}) - \gamma\boldsymbol{\psi} - i\mathbf{F}(t), \quad (\text{S24})$$

i.e. for $m \in \{A, B, L, R\}$,

$$\frac{d\psi_m}{dt} = -i \sum_n H_{mn} \psi_n - iU |\psi_m|^2 \psi_m - \gamma \psi_m - iF_m(t). \quad (\text{S25})$$

Here H is Hermitian, γ is a uniform loss rate, U is a Kerr nonlinearity, and $\mathbf{F}(t)$ drives only the inputs. In all comparator simulations we use the linear regime $U = 0$. Instantaneous readout intensities are $I_{L,R}(t) = |\psi_{L,R}(t)|^2$.

S4.2 Four-mode junction Hamiltonian

We use

$$H = J \begin{pmatrix} +\Delta/J & 0 & +1 & -1 \\ 0 & -\Delta/J & +1 & +1 \\ +1 & +1 & +\omega_0/J & \kappa/J \\ -1 & +1 & \kappa/J & -\omega_0/J \end{pmatrix}, \quad (\text{S26})$$

with

$$J = 1, \quad \Delta = 1.20 J, \quad \omega_0 = 1.20 J, \quad \kappa = 0.50 J. \quad (\text{S27})$$

The signed input–output couplings implement a flux-biased interferometer, while the opposite detunings ($\pm\Delta$ on A, B and $\pm\omega_0$ on L, R) together with readout hybridization κ convert order-dependent interference into a robust left/right energy imbalance.

S4.3 Pulse drives

Only the input modes are driven,

$$\mathbf{F}(t) = (F_A(t), F_B(t), 0, 0)^\top, \quad (\text{S28})$$

with Gaussian pulses

$$F_A(t) = P_0 \exp\left[-\frac{(t-t_A)^2}{2\sigma^2}\right], \quad F_B(t) = P_0 \exp\left[-\frac{(t-t_B)^2}{2\sigma^2}\right]. \quad (\text{S29})$$

For Fig. 4 we use

$$P_0 = 6.0, \quad \sigma = 0.25 J^{-1}, \quad \delta t \equiv |t_A - t_B| = 1.60 J^{-1}, \quad (\text{S30})$$

with $t_{\text{ref}} = 5.0$ and the two orderings

$$\begin{aligned} \text{(a)} \quad & t_A = t_{\text{ref}}, \quad t_B = t_{\text{ref}} + \delta t, \\ \text{(b)} \quad & t_A = t_{\text{ref}} + \delta t, \quad t_B = t_{\text{ref}}. \end{aligned} \quad (\text{S31})$$

S4.4 Initialization and integration

We initialize at rest, $\psi(0) = \mathbf{0}$. Equations (S24)–(S25) are integrated with `solve_ivp` (adaptive RK45) over $t \in [0, T_{\text{final}}]$ with $T_{\text{final}} = 25.0 J^{-1}$ and sampled on a uniform grid

$$t_{\text{eval}} = 0, dt, 2dt, \dots, T_{\text{final}}, \quad dt = 0.005 J^{-1}, \quad (\text{S32})$$

using tolerances $\text{rtol} = 10^{-6}$ and $\text{atol} = 10^{-8}$.

S4.5 Integrated-energy readout

To suppress transient beating, the decision uses time-integrated readout energies over a window anchored to the last pulse. Let $t_{\text{last}} = \max(t_A, t_B)$ and define

$$t_0 = t_{\text{last}} + 2\sigma, \quad T_w = 6\sigma. \quad (\text{S33})$$

The energies are

$$E_L = \int_{t_0}^{t_0+T_w} |\psi_L(t)|^2 dt, \quad E_R = \int_{t_0}^{t_0+T_w} |\psi_R(t)|^2 dt, \quad (\text{S34})$$

computed by trapezoidal integration on the sampled trajectory (S32). We report

$$b_{AB} = \mathbb{I}[E_L > E_R], \quad m_{AB} = \log\left(\frac{E_L}{E_R}\right), \quad (\text{S35})$$

where $m_{AB} > 0$ corresponds to routing left (interpreted as $t_A < t_B$ under the convention of H), and $m_{AB} < 0$ to routing right.

S4.6 Three-spike permutation decoding and cycle repair (Fig. 5)

Figure 5 composes three independent instances of the same $N = 2$ comparator for pairs (A, B) , (B, C) , and (A, C) , followed by a deterministic repair rule applied to the comparator margins.

Trial generation. For each target permutation of (A, B, C) , generate ordered times

$$(t_1, t_2, t_3) = (t_{\text{ref}}, t_{\text{ref}} + g_1, t_{\text{ref}} + g_1 + g_2), \quad (\text{S36})$$

with $t_{\text{ref}} = 5.0$ and gaps

$$g_{1,2} = \Delta t [1 + 0.10 \mathcal{N}(0, 1)], \quad \Delta t = 1.60 J^{-1}, \quad (\text{S37})$$

subject to $\text{min_sep} = 0.4 J^{-1}$. Assign (t_1, t_2, t_3) to (t_A, t_B, t_C) according to the target permutation.

Pairwise comparator outputs. For each ordered pair $(X, Y) \in \{(A, B), (B, C), (A, C)\}$, run an independent comparator and compute

$$b_{XY} = \mathbb{I}[E_L^{(XY)} > E_R^{(XY)}], \quad m_{XY} = \log\left(\frac{E_L^{(XY)} + \epsilon}{E_R^{(XY)} + \epsilon}\right), \quad (\text{S38})$$

with $\epsilon = 10^{-12}$ to regularize rare near-ties. The bit $b_{XY} = 1$ corresponds to $t_X < t_Y$ under the sign convention of Eq. (S26).

Tournament code and repair. The three bits (b_{AB}, b_{BC}, b_{AC}) form a tournament code. If the code is cyclic, repair by flipping the weakest comparison:

$$(X^*, Y^*) = \arg \min\{|m_{AB}|, |m_{BC}|, |m_{AC}|\}, \quad b_{X^*Y^*} \leftarrow 1 - b_{X^*Y^*}. \quad (\text{S39})$$

Decode the (repaired) transitive code to a permutation using the standard mapping. We generate 200 trials per permutation (seed 42). Figure 5(a) reports the column-normalized confusion matrix, and Fig. 5(b) shows the distribution of

$$|m|_{\min} = \min(|m_{AB}|, |m_{BC}|, |m_{AC}|), \quad (\text{S40})$$

separately for acyclic and cyclic codes before repair.

S5 Calibration and hardware-in-the-loop optimization

This section details the calibration setting used in Sec. VII and Fig. 6. We treat the device as a black box and optimize programmable controls using only observable readouts (integrated output energies). We specify the static disorder model and the SPSA-based hardware-in-the-loop (HIL) protocol used to recover the $N = 2$ comparator under strong mismatch.

S5.1 Static disorder model (Fig. 6)

We model fabrication mismatch as fixed, unknown perturbations to coupling phases, coupling magnitudes, and detunings. All disorder is static across trials and is not revealed to the calibration algorithm.

Phase offsets and controls. The effective programmable phases are

$$\boldsymbol{\theta}^{\text{eff}} = \boldsymbol{\theta}^{\text{fab}} + \boldsymbol{\theta}^{\text{ctrl}}, \quad (\text{S41})$$

with unknown device offsets $\boldsymbol{\theta}^{\text{fab}} \in \mathbb{R}^4$ and optimized controls $\boldsymbol{\theta}^{\text{ctrl}} \in \mathbb{R}^4$. In the Fig. 6 ensemble,

$$\theta_p^{\text{fab}} \sim \mathcal{N}(0, \sigma_{\text{fab}}^2), \quad \sigma_{\text{fab}} = 1.5 \text{ rad}, \quad p = 1, \dots, 4, \quad (\text{S42})$$

for the four input→output couplings (AL, AR, BL, BR). The specific instance in Fig. 6 uses

$$\boldsymbol{\theta}^{\text{fab}} = [0.002, 0.448, -0.411, -1.336] \text{ rad}. \quad (\text{S43})$$

Amplitude imbalance (lognormal, clipped). We perturb the four coupling magnitudes by multiplicative factors $s_p > 0$,

$$g_p \rightarrow g_p s_p, \quad p \in \{AL, AR, BL, BR\}, \quad (\text{S44})$$

drawing s_p i.i.d. from a lognormal distribution with 25% RMS and clipping to $s_p \in [0.5, 1.5]$. For Fig. 6,

$$[s_{AL}, s_{AR}, s_{BL}, s_{BR}] = [1.015, 1.335, 0.877, 0.845]. \quad (\text{S45})$$

Detuning disorder. We apply static detuning shifts

$$\Delta \rightarrow \Delta + \delta\Delta, \quad \omega_0 \rightarrow \omega_0 + \delta\omega_0, \quad \delta\Delta, \delta\omega_0 \sim \mathcal{N}(0, \sigma_{\text{det}}^2), \quad (\text{S46})$$

with $\sigma_{\text{det}} = 0.15 J$. For the Fig. 6 instance,

$$\delta\Delta = -0.0682 J, \quad \delta\omega_0 = -0.1487 J. \quad (\text{S47})$$

Disordered, programmable junction Hamiltonian. We integrate the four-mode equations of Sec. S4 with $U = 0$, but replace the ideal input→readout couplings by

$$\begin{aligned} g_{AL} &= +J s_{AL} e^{i\theta_{AL}^{\text{eff}}}, & g_{AR} &= -J s_{AR} e^{i\theta_{AR}^{\text{eff}}}, \\ g_{BL} &= +J s_{BL} e^{i\theta_{BL}^{\text{eff}}}, & g_{BR} &= +J s_{BR} e^{i\theta_{BR}^{\text{eff}}}, \end{aligned} \quad (\text{S48})$$

with $\theta_p^{\text{eff}} = \theta_p^{\text{fab}} + \theta_p^{\text{ctrl}}$. The resulting Hermitian Hamiltonian is

$$H(\boldsymbol{\theta}^{\text{eff}}, s) = \begin{pmatrix} +\Delta & 0 & g_{AL} & g_{AR} \\ 0 & -\Delta & g_{BL} & g_{BR} \\ g_{AL}^* & g_{BL}^* & +\omega_0 & \kappa \\ g_{AR}^* & g_{BR}^* & \kappa & -\omega_0 \end{pmatrix}. \quad (\text{S49})$$

S5.2 Comparator readout

Each trial uses the same Gaussian pulse drives and integrated-energy readout window as in Sec. S4. Let $I_{L,R}(t)$ be instantaneous intensities and define

$$E_L = \int_{t_0}^{t_0+T_w} I_L(t) dt, \quad E_R = \int_{t_0}^{t_0+T_w} I_R(t) dt, \quad (\text{S50})$$

with $t_0 = t_{\text{last}} + 2\sigma$ and $T_w = 6\sigma$. The hard decision is $\arg \max[E_L, E_R]$, and we also use the signed log-margin

$$m_{AB} = \log\left(\frac{E_L}{E_R}\right). \quad (\text{S51})$$

S5.3 Calibration objective

We optimize $\boldsymbol{\theta}^{\text{ctrl}}$ using only measured (E_L, E_R) . For a labeled trial with $y \in \{L, R\}$, define a softmax over energies

$$p_k = \frac{\exp(\beta E_k)}{\exp(\beta E_L) + \exp(\beta E_R)}, \quad k \in \{L, R\}, \quad \beta = 1, \quad (\text{S52})$$

and per-sample cross-entropy $\ell = -\log p_y$. For a batch of size B we use

$$\mathcal{L} = \frac{1}{B} \sum_{i=1}^B \ell_i + \lambda \|\boldsymbol{\theta}^{\text{ctrl}}\|_2^2, \quad \lambda = 10^{-4}. \quad (\text{S53})$$

S5.4 SPSA update rule

At step k , draw $\boldsymbol{\Delta} \in \{\pm 1\}^4$ i.i.d. and use constant step sizes

$$a_k = \xi, \quad c_k = \delta, \quad \xi = 0.08, \quad \delta = 0.08 \text{ rad}, \quad (\text{S54})$$

over 250 steps. Evaluate $\mathcal{L}^\pm = \mathcal{L}(\boldsymbol{\theta}^{\text{ctrl}} \pm c_k \boldsymbol{\Delta})$ on the current batch, form

$$\widehat{\mathbf{g}} = \frac{\mathcal{L}^+ - \mathcal{L}^-}{2c_k} \boldsymbol{\Delta}, \quad (\text{S55})$$

and update

$$\boldsymbol{\theta}^{\text{ctrl}} \leftarrow \boldsymbol{\theta}^{\text{ctrl}} - a_k \widehat{\mathbf{g}}, \quad (\text{S56})$$

wrapping phases to $[0, 2\pi)$ after each update.

S5.5 Training and evaluation protocol (Fig. 6)

Training batches. Each SPSA step uses $B = 48$ examples with balanced labels. For each example, set $t_{\text{ref}} = 5.0/J$, sample a separation

$$\text{sep} \sim \delta t (1 + 0.30 \mathcal{N}(0, 1)), \quad \delta t = 1.60/J,$$

clip $\text{sep} \leftarrow \max(\text{sep}, 0.8\sigma)$, add timing jitter with $\sigma_t = 0.15\sigma$ to t_A, t_B , and assign (t_A, t_B) according to the label.

Fixed primitive parameters. We use $U = 0$, $J = 1$, $\gamma = 0.10J$, $\Delta_0 = \omega_{0,0} = 1.20J$, $\kappa = 0.50J$, $P_0 = 6.0$, $\sigma = 0.25/J$, and integrate to $T_{\text{final}} = 25/J$ (as in Sec. S4).

Learned correction (Fig. 6 instance). For the disorder instance above, SPSA learns

$$\boldsymbol{\theta}^{\text{ctrl}} = [-0.065, -1.236, -0.012, +0.802] \text{ rad.} \quad (\text{S57})$$

Evaluation. We evaluate before/after calibration on 800 held-out trials (400 per class) and report the column-normalized confusion matrix. Before calibration ($\boldsymbol{\theta}^{\text{ctrl}} = \mathbf{0}$),

$$\text{CM}_{\text{before}} = \begin{pmatrix} 0.995 & 0.878 \\ 0.005 & 0.122 \end{pmatrix}, \quad (\text{S58})$$

(mean diagonal 55.9%), while after calibration,

$$\text{CM}_{\text{after}} = \begin{pmatrix} 1.000 & 0.055 \\ 0.000 & 0.945 \end{pmatrix}, \quad (\text{S59})$$

(mean diagonal 97.2%). Figure 6 also reports the batch loss, batch accuracy, and median $|m_{AB}|$ versus step.

S5.6 Measurement noise and limits

Calibration compensates static mismatch (fixed phase offsets, coupling imbalance, and detuning shifts) but cannot remove trial-to-trial timing jitter. With noisy measurements of $E_{L,R}$, SPSA remains applicable but may require averaging to reduce the variance of \mathcal{L}^{\pm} .

S6 End-to-end MoE routing benchmark: numerical details

This section documents the benchmark used for Fig. 7 (Sec. VIII D). The goal is to map a routing-error rate at the device level to task accuracy in a sparse top-1 routing setting, while keeping expert computation fixed.

S6.1 Setup: hard top-1 routing with injected errors

We consider a minimal mixture-of-experts (MoE) classifier with hard top-1 routing. Each input is associated with a ground-truth expert label $k^* \in \{1, \dots, K\}$; the router outputs \hat{k} . We inject routing errors independently of expert models and control

$$p_{\text{route}} \equiv \Pr(\hat{k} \neq k^*). \quad (\text{S60})$$

The benchmark reports task accuracy as a function of p_{route} .

S6.2 Synthetic data and labels

Inputs are vectors $x \in \mathbb{R}^{10}$ drawn from a Gaussian mixture. For each random seed, sample cluster means $\{\mu_k\}_{k=1}^K$ i.i.d. from a standard normal and hold them fixed. For each sample, draw k^* uniformly from $\{1, \dots, K\}$ and then

$$x \sim \mathcal{N}(\mu_{k^*}, \sigma_x^2 I). \quad (\text{S61})$$

Table 3: Representative device operating points used in Fig. 7.

Point	σ_t/T_{wrap}	σ_θ (rad)	P_{correct}	p_{route}	Regime
A	0.00	0.0	1.000	0.000	supported
B	0.02	0.1	1.000	0.000	supported
C	0.05	0.2	1.000	0.000	supported
D	0.10	0.3	0.998	0.002	supported
E	–	–	–	0.08	stressed (projected)
F	–	–	–	0.15	stressed (projected)

Binary labels are generated by an expert-specific linear rule,

$$y = \mathbb{1} \left[\mathbf{w}_{k^*}^\top x + b_{k^*} > 0 \right], \tag{S62}$$

with \mathbf{w}_k, b_k sampled at initialization and held fixed per seed. Two difficulty settings are used: $\sigma_x = 0.6$ (well-separated clusters) and $\sigma_x = 0.9$ (moderate overlap).

S6.3 Experts and training

Each expert is a logistic-regression classifier trained only on samples from its own cluster ($k^* = k$), using batch gradient descent with early stopping. We report $K \in \{8, 32\}$ and average results over 10 random seeds.

S6.4 Routing-error injection

Routing is modeled as

$$\hat{k} = \begin{cases} k^*, & \text{with probability } 1 - p_{\text{route}}, \\ \text{Uniform}(\{1, \dots, K\} \setminus \{k^*\}), & \text{with probability } p_{\text{route}}. \end{cases} \tag{S63}$$

For each $p_{\text{route}} \in [0, 0.5]$, accuracy is evaluated on 5,000 held-out test samples. For $K > 2$, uniform random top-1 routing corresponds to $p_{\text{route}} = 1 - 1/K$.

S6.5 Device-to-benchmark mapping

We map device address-selection accuracy to routing error via

$$p_{\text{route}} = 1 - P_{\text{correct}}. \tag{S64}$$

Supported operating points use P_{correct} extracted from the WTA simulations in Fig. 3 (main text), while two additional stressed points (E–F) illustrate higher error rates outside the simulated grid.

The operating points used in Fig. 7 are listed in Table 3.

S6.6 Evaluation scope and cascading

For each $(K, \sigma_x, p_{\text{route}})$ configuration, accuracy is averaged over 10 seeds; Fig. 7 shows the mean with a $\pm 1\sigma$ band. The benchmark is intentionally routing-limited: it does not model expert capacity limits, load balancing, soft routing, or end-to-end training.

Figure 7 characterizes a single routing event. For a cascade of L stages with independent routing error p_{route} , the probability of fully correct routing is $(1 - p_{\text{route}})^L$. In a hybrid implementation,

Table 4: **Dominant scaling of the address-selection step (order-of-growth).** Here $P \sim NK$ denotes the number of programmable controls in the interferometric map, and n_c denotes the number of primitive operations required to score K candidates against N spikes in a conventional implementation (typically $n_c = \mathcal{O}(NK)$, plus an $\mathcal{O}(K)$ reduction for an argmax). Latencies are shown as time complexity at fixed resources; “hardware” indicates spatial/parallel resources.

	Conventional compare-select	digitize-	Polychronous wave score- select
Front-end representation	timestamp/bin/rate extraction		phase-referenced event encoding
Score evaluation (time)	$\mathcal{O}(n_c)$ (logic and bandwidth)		$\mathcal{O}(1)$ in time (parallel interference)
Score evaluation (hardware)	modest compute, memory traffic		interferometer with $\mathcal{O}(P)$ programmable controls
Selection (time)	$\mathcal{O}(K)$ (argmax / reduction tree)		$\mathcal{O}(1)$ (WTA settling)
Selection (hardware)	comparator tree / digital reduction		competitive modes + gain saturation
Calibration overhead	standard mixed-signal calibration		phase tuning with intensity-only feedback (Sec. VII)

the winning mode is detected to produce the discrete index \hat{k} , and conventional control programs the next stage’s inputs; this control plane operates on the low-rate address and need not resolve ultrafast spike timing.

Figure 7 quantifies the system-level impact of imperfect address selection by explicitly injecting routing errors into a hard top-1 mixture-of-experts benchmark while holding expert computation fixed. The results show that task accuracy degrades smoothly with routing error probability, with no abrupt failure, and that the routing error rates predicted by the driven-dissipative device simulations correspond to negligible performance loss in both specialized and moderately overlapping expert regimes. In this benchmark, the predicted routing error rates from Fig. 3 fall in a regime where task accuracy drops only slightly. This does not guarantee the same tolerance for all workloads, but it provides a concrete mapping from measured routing error to downstream performance.

S7 End-to-end scaling snapshot (interface-limited baseline vs. wave-domain selection).

Here we summarize the dominant scaling of the address-selection step for a decision window with N active spikes and K candidate templates (or experts). Table 4 contrasts a conventional digitize-compare-select path with polychronous wave scoring followed by WTA digitization.