Loops as States

Mapping Superconducting Optoelectronic Loop Dynamics
onto Diagonal State-Space Sequence Models

Micheal Ohagwu

Nani Research

obi@nani-inc.com | New York

June 16, 2026

Technical note (v2) — comments welcome

Abstract

State-space sequence models (S4, S4D, Mamba) are a leading efficient alternative to attention; superconducting optoelectronic networks (SOENs) are a neuromorphic hardware program whose synapses, dendrites, and somas each obey one first-order "leaky-integrator" equation. We make the correspondence between them explicit: *in its linear regime a SOEN loop is a single diagonal state-space mode.* A leaky loop is a real pole; a resonant (second-order) dendrite is a complex-conjugate pair—the oscillatory modes S4D initializes—and the source-function nonlinearity supplies both the inter-block nonlinearity and a physical analog of Mamba's input-dependent timescales. The linearized SOEN element and a diagonal SSM are the same continuous-time system and share the same discrete recurrence under matched discretization. We provide an explicit dictionary (Table 1) and a controlled realization experiment: mapping a target S4D layer onto perturbed SOEN loops is exact at zero jitter, tolerates $\sim 14\%$ variability in decay/timescale, but only $\sim 0.45\%$ in mode frequency—phase accumulation makes frequency precision, not leak precision, the binding constraint—while state noise degrades reconstruction gracefully. At the task level (fine frequency discrimination on a fixed resonator bank with a trained read-out) the same frequency sensitivity appears but is softer, and device-aware training—fitting the read-out to jitter-augmented features—recovers much of the lost accuracy without touching the hardware. Prior work has placed SSMs on digital-neuromorphic, memristive, and op-amp substrates; our specific contribution is the SOEN loop-physics-to-diagonal-SSM dictionary together with this tolerance analysis. We are explicit throughout about the linearization and hardware assumptions the mapping rests on.

1 Motivation

Two research programs have converged on the same mathematical object from opposite directions. From machine learning, *structured state-space models* (SSMs) replaced the quadratic-cost attention mechanism for long sequences with a bank of linear dynamical systems: S4 [1], its diagonal simplification S4D [2], the linear recurrent unit (LRU) [4], and the selective model Mamba [3]. From hardware, *superconducting optoelectronic networks* (SOENs) [6, 7] build brain-inspired circuits whose every active element—synapse, dendrite, soma—reduces, in the phenomenological model derived from a circuit Lagrangian, to one first-order ordinary differential equation. The central observation of this note is that these are, in the linear regime, *the same equation.* Making the dictionary explicit clarifies (a) what SOEN hardware computes in the language ML practitioners already use, and (b) why an energy-based neuromorphic substrate is a natural physical home for SSM inference.

Contributions. (1) An explicit, checkable dictionary between SOEN circuit dynamics and diagonal-SSM components (§3, Table 1). (2) A discretization argument showing the SOEN phenomenological update and the S4D recurrence coincide exactly (continuous-time / ZOH) and to first order (Euler), with real poles ↔ leaky integrators and complex poles ↔ resonant dendrites (§4). (3) A reading of input-dependent source functions as selective (Mamba-style) timescales (§5). (4) An itemized energy account and a substrate-aware design stance (§6). (5) A controlled realization experiment isolating mode frequency as the binding constraint (§7). (6) A task-level study showing the constraint persists but softens for magnitude read-outs, and that device-aware training recovers much of the lost accuracy (§8). We state limitations plainly in §9.

Relation to prior work. Mapping SSMs onto efficient hardware is now an active area, and we position against it rather than claim it. S4D has been deployed on Intel's Loihi 2 digital-neuromorphic processor with large streaming energy/latency gains [9]; S4/S4D kernels have been placed on memristive in-memory crossbars [10] and other compute-in-memory substrates for event-sequence processing [11]; and diagonal SSMs have been synthesized directly into analog operational-amplifier low-pass circuits [12], the closest classical-electronics analog of the leaky-loop argument here. The continuous-time-system view of sequence models traces to the linear state-space layer [8]. Against this backdrop our contribution is narrow and specific: a dictionary from *superconducting optoelectronic loop physics* (single-photon communication, flux-state memory, resonant dendrites) to diagonal SSM modes, plus a tolerance analysis that names frequency precision as the first-order bottleneck. We do not claim the first hardware SSM, nor any measured device.

2 Background in one page

The SOEN element. In the phenomenological model [7], a SOEN dendrite/soma holds a dimensionless loop signal $s = I/I_c$ driven by applied flux ϕ from upstream loops:

$$\tau\,\dot{s} = -\,s \;+\; \gamma\,\tau\,g(\phi, s), \qquad \phi = \textstyle\sum_j J_j\, s_j, \tag{1}$$

where g is the SQUID *source function* (a steep, saturating nonlinearity set by a bias current i_b), γ a gain, τ a leak time, and J_j flux-coupling weights. Equation (1) is a leaky integrator; the full circuit dendrite (a SQUID feeding an LRC branch) is in general a damped, driven second-order oscillator, capable of resonances and subthreshold oscillations [7].

The state-space model. An SSM layer maps an input sequence $u(t) \in \mathbb{R}$ to an output $y(t)$ through a latent state $x(t) \in \mathbb{C}^N$:

$$\dot{x} = A\,x + B\,u, \qquad y = C\,x\;(+\,D\,u). \tag{2}$$

S4D takes $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ diagonal with $\mathrm{Re}\,\lambda_n < 0$ [2]. Discretizing with step Δ (zero-order hold) gives the recurrence and convolution

$$x_k = \bar{A}\,x_{k-1} + \bar{B}\,u_k, \quad y_k = C\,x_k, \qquad \bar{A} = e^{\Delta A}, \quad y = u * K, \;\; K_\ell = C\bar{A}^\ell\bar{B}. \tag{3}$$

For diagonal A the kernel is a sum of geometric (i.e. sampled exponential) sequences, $K_\ell = \sum_n C_n\,\bar{\lambda}_n^\ell\,\bar{B}_n$, with $\bar{\lambda}_n = e^{\Delta\lambda_n}$.

3 The dictionary

We linearize (1) about an operating point (ϕ^*, s^*). Writing $g \approx g^* + g_\phi\,(\phi - \phi^*) + g_s\,(s - s^*)$ and absorbing constants into the input, the deviation $x \equiv s - s^*$ obeys

$$\dot{x} = -\,\alpha\,x + b\,u, \qquad \alpha = \frac{1 - \gamma\tau g_s}{\tau}, \quad b = \gamma g_\phi, \quad u \equiv \phi - \phi^*. \tag{4}$$

This is exactly (2) for a single state with $A = -\alpha$, a real pole. A *layer* of such loops with diverse leak times $\{\tau_n\}$ and biases is a diagonal SSM with real modes $\lambda_n = -\alpha_n$.

Scope of "diagonal." The mapping is diagonal precisely when the loops are driven in parallel by a common input and combined only by the read-out C. With recurrent flux coupling $\phi_i = \sum_j J_{ij} s_j$ between *dynamic* loops, the coupling enters the state matrix and A is no longer diagonal: the layer is then a general linear state-space model (equivalently a linear RNN), of which diagonal S4D is the special case. The diagonal dictionary below should be read as the parallel-bank case; recurrent SOEN topologies realize the broader dense-A family.

Resonant dendrites give complex poles. A pure leaky loop has a single real pole. Restoring the dendrite's second-order LRC branch, the linearized signal obeys a damped driven oscillator,

$$\ddot{x} + 2\alpha\,\dot{x} + (\alpha^2 + \omega^2)\,x = b\,u, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -(\alpha^2+\omega^2) & -2\alpha \end{pmatrix}\begin{pmatrix} x \\ \dot{x} \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix} u, \tag{5}$$

whose characteristic polynomial $\lambda^2 + 2\alpha\lambda + (\alpha^2+\omega^2) = 0$ has roots $\lambda_\pm = -\alpha \pm i\omega$. Diagonalizing the real 2×2 block gives the complex-conjugate pair $-\alpha \pm i\omega$: one S4D complex mode. Setting $\omega \to 0$ collapses the pair to a repeated (critically damped) real pole at $-\alpha$; the genuine first-order leaky integrator of (4) is the overdamped regime, where the second-order branch has two distinct real poles and its fast pole is dropped. Table 1 is the resulting correspondence.

SOEN (hardware)	**Diagonal SSM (S4D / Mamba)**
loop signal $s = I/I_c$	state coordinate x_n
applied flux $\phi = \sum_j J_{ij} s_j$	external drive Bu (parallel bank); off-diagonal A (recurrent)
leak time τ_n	mode timescale; pole $\lambda_n = -1/\tau_n$
resonant LRC dendrite (subthreshold osc.)	complex mode $\lambda_n = -\alpha_n \pm i\omega_n$
spread of τ_n across dendrites	HiPPO / S4D timescale initialization
explicit update $s \leftarrow (1 - \frac{\Delta}{\tau})s + \frac{\Delta}{\tau}u$	discrete scan $x_k = \bar{A}x_{k-1} + \bar{B}u_k$
dendritic impulse response $\propto e^{-t/\tau}$	SSM convolution kernel $K_\ell \propto \bar{\lambda}^\ell$
source function g, soma threshold	pointwise nonlinearity between SSM blocks
flux/bias-dependent effective $\tau(\phi)$	selective (input-dependent) Δ, A of Mamba
attojoule-scale loop relaxation	one cheap state-update step

Table 1: An explicit dictionary between superconducting optoelectronic loop dynamics and diagonal state-space sequence models. The correspondence is exact in the linear regime (4) and approximate once the source-function nonlinearity is restored.

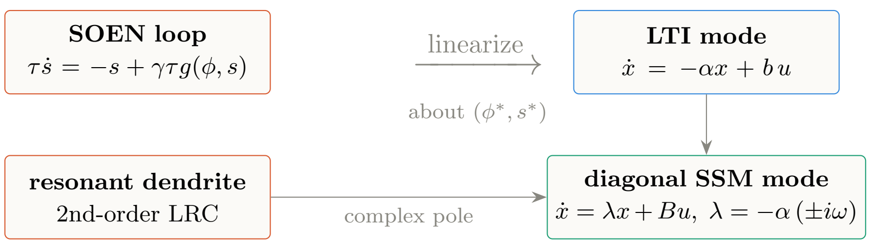

Figure 1: One SOEN loop, linearized, is one diagonal SSM mode. A pure leaky integrator yields a real pole; a resonant (second-order LRC) dendrite yields a complex-conjugate pair—the oscillatory modes S4D is built to represent.

4 Equivalence by discretization

It is worth being precise about *which* equivalence holds, since "the same recurrence" can mean three different things.

1. **Continuous time (exact).** The linearized loop (4) and a diagonal SSM mode (2) are the *same* LTI system, $\dot{x} = -\alpha x + bu$ with $A = -\alpha$.
2. **Exact (ZOH) discretization.** Sampling (4) with step Δ under zero-order hold gives $x_k = e^{-\alpha\Delta} x_{k-1} + \bar{b}\, u_k$, $\bar{b} = (e^{-\alpha\Delta} - 1)\, b/(-\alpha)$ — exactly the S4D recurrence (3) with $\bar{A} = e^{\Delta A}$.
3. **Explicit Euler (first order).** The forward-Euler update $x_k = (1 - \alpha\Delta) x_{k-1} + b\Delta\, u_k$ — equivalently $s \leftarrow s + (\gamma g - s/\tau)\Delta$ — is what SOEN phenomenological simulators and our interactive models actually run. It matches the ZOH recurrence to first order, $1 - \alpha\Delta = e^{-\alpha\Delta} + O(\Delta^2)$.

So SOEN and S4D coincide *exactly* in continuous time and under matched (ZOH) discretization, and to first order in Δ when the device is Euler-simulated while S4D uses ZOH. Either way, unrolling the recurrence gives the output as a causal convolution with a (sampled) exponential kernel,

$$y_k = \sum_{\ell \geq 0} K_\ell\, u_{k-\ell}, \qquad K_\ell = C\, \bar{A}^\ell\, \bar{B} \ \propto\ e^{-\alpha\Delta\,\ell}, \tag{6}$$

i.e. the dendrite's exponential impulse response *is* a single SSM convolution kernel (Fig. 2a). Stacking modes with different poles builds an arbitrary multi-exponential (and, with complex poles, damped-oscillatory) kernel—the S4D basis (Fig. 2b).

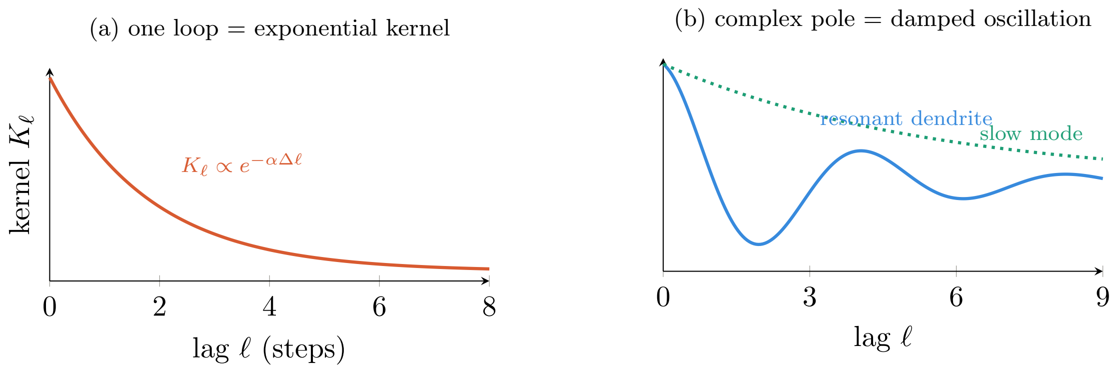

Figure 2: (a) A single leaky loop contributes an exponential convolution kernel. (b) A resonant (complex-pole) dendrite contributes a damped oscillation; a bank of loops with varied (α_n, ω_n) spans the same kernel family S4D parameterizes.

5 Selectivity: the nonlinear part

The linear dictionary above captures S4/S4D exactly. Mamba's advance is *selectivity*: making Δ, A, B functions of the input so the model can gate and choose timescales per token [3]. SOEN supplies a physical version of the same idea. Because the source function $g(\phi, s)$ is nonlinear and flux-dependent, the *effective* decay $\alpha(\phi) = \big(1 - \gamma\tau\, g_s(\phi)\big)/\tau$ and gain depend on the instantaneous drive: strong flux changes how fast the loop integrates and forgets. Read in SSM terms, the device implements an input-dependent timescale—a hardware analog of selective state spaces—without any explicit gating network, at the cost of being analog and approximate rather than learned and exact. We state this as an analogy to be tested, not an identity: quantifying how closely a biased SOEN dendrite reproduces a trained selective $\Delta(u)$ is concrete future work.

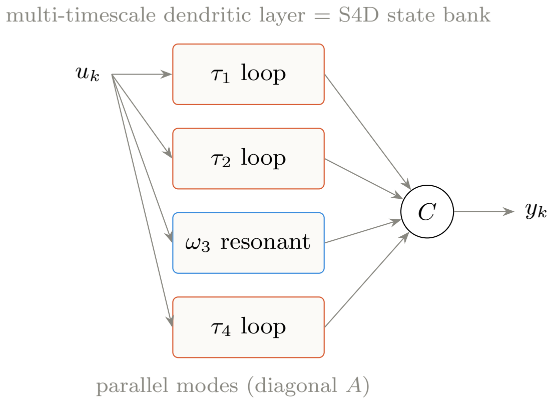

Figure 3: A layer of independent SOEN loops with diverse timescales is a diagonal state-space layer: each loop is one state coordinate, the read-out C mixes them, and the source-function nonlinearity (omitted here) plays the role of the pointwise nonlinearity between SSM blocks.

6 The energy corollary

It is more honest to itemize the energy than to quote a single number. Per state-update step the relevant terms are: (i) the *synaptic detection / loop update*, a single-photon event at the $\sim 10^{-17}$ J (tens of aJ) scale [6]; (ii) the *optical emission* when a soma fires, set by photons-per-spike and source (laser/LED) wall-plug efficiency—the dominant on-chip term and the least mature; (iii) *read-out and control*—amplification, clocking, converters at the cryostat boundary; and (iv) the *cooling overhead*, $\sim 10^3$ warm watts per cold watt at 4 K, amortized only by high utilization and fan-out. The dictionary makes the accounting concrete: an N-state SSM layer is N loop updates of type (i)–(ii) per step, so the recurrence is already what the hardware does. But the headline "attojoule" figure refers to term (i) alone; a fair comparison against a CMOS multiply–accumulate must include (ii)–(iv). We therefore claim a low per-update *floor*, not a system-level win, and flag (ii) as the quantity most in need of measurement.

One-line summary. A SOEN loop, linearized and discretized, is a diagonal state-space mode; a multi-timescale dendritic layer is an S4D layer; the source-function nonlinearity supplies both the inter-block nonlinearity and a physical analog of Mamba's selective timescales—so SSMs are the natural "software twin" of superconducting loop hardware.

7 A realization experiment

We test the dictionary numerically rather than rhetorically: take a target diagonal S4D layer, realize its modes as SOEN loops with realistic non-idealities, and measure how far the realized output drifts from the ideal.

Setup. The target is $N = 8$ complex modes in the S4D-Lin family, $\lambda_n = -\frac{1}{2} + i\pi n$, with a fixed read-out C, zero-order-hold step $\Delta = 0.05$, and sequence length $L = 400$ (window $T = 20$). The realization uses the *same* modes but perturbs each device: *fabrication jitter* scales the decay α_n and frequency ω_n by $(1 + \sigma\varepsilon)$, $\varepsilon \sim \mathcal{N}(0, 1)$, drawn independently per mode and per device; and additive *state noise* σ_s (thermal/flux) is injected each step, scaled to the signal. The error is the normalized RMSE of the realized output against the ideal S4D output on held-out white-noise test inputs, averaged over Monte-Carlo device realizations.

Results. Four findings (Figs. 4, 5):

1. At zero perturbation the NRMSE is 0: the linear dictionary is *exact*, as §4 predicts.
2. **Decay/timescale jitter is benign**: NRMSE stays under 0.1 up to $\sigma_\alpha \approx 14\%$. The dissipative envelope is forgiving, so leak times and biases may vary substantially (Fig. 5a—the realization still tracks the target).
3. **Frequency jitter is the binding constraint**: a fine sweep ($M = 120$, ±SEM; Fig. 6) places the NRMSE= 0.1 crossing at $\sigma_\omega \approx 0.45\%$, ∼30× tighter than the decay tolerance. A small relative frequency error integrates into a large phase error over the window (phase $\sim \sigma_\omega \, \omega T$), so the waveform decorrelates while keeping the right envelope (Fig. 5b). The crossing is within a factor ~ 2 of the $1/(\omega_{\max} T) \approx 0.2\%$ phase-accumulation estimate (the gap reflects averaging over modes and the 0.1 threshold).
4. **State noise degrades gracefully**, roughly linearly, reaching NRMSE 0.1 near $\sigma_s \approx 2\%$ of signal.

Takeaway. Realizing oscillatory S4D modes on analog SOEN hardware is a *frequency-precision* problem, not a timescale-precision problem: leak times can drift by $\sim$ 10%, but resonant-dendrite frequencies must be trimmed to $\lesssim 1\%$. This is an actionable design constraint—favor real-pole (overdamped) modes where the task allows, and add frequency trimming or locking to resonant elements—and it is exactly the kind of quantitative guidance the bare dictionary cannot give.

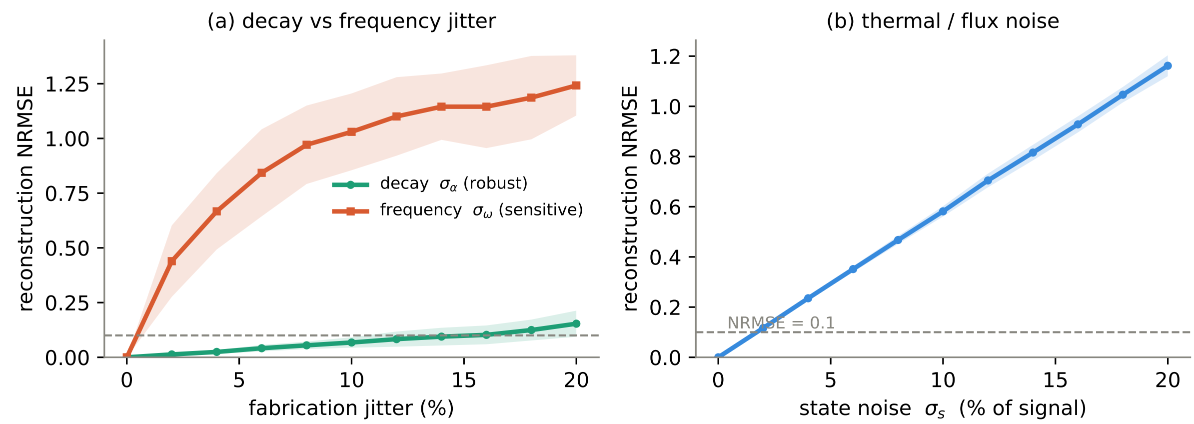

Figure 4: Reconstruction error of a SOEN realization versus an ideal S4D layer. (a) Decay-timescale jitter σ_α (teal) is tolerated to $\sim$ 14%, while frequency jitter σ_ω (coral) crosses NRMSE = 0.1 near 0.5%. (b) Additive state (thermal/flux) noise degrades reconstruction gracefully and roughly linearly. Bands are ±1 s.d. over device realizations.

8 Task-level realization and device-aware training

Kernel NRMSE is phase-sensitive, so it is the harshest possible metric. Two questions remain: does the frequency-precision constraint survive at the level of a *task*, and can it be mitigated? On an analog substrate one typically cannot backpropagate through the device, but one *can* train a read-out; we therefore adopt a realistic, deployment-inspired setup: a *fixed* resonator bank (a diagonal SSM used as a feature extractor) with a trained linear read-out — a reservoir-style model. We treat the following as a controlled toy demonstration, not validation of general hardware SSM inference.

Setup. A 3-class fine frequency-discrimination task: inputs are noisy sinusoids whose frequency falls in one of three closely-spaced bands ($\omega \in \{1.00, 1.05, 1.10\}$, spacing comparable

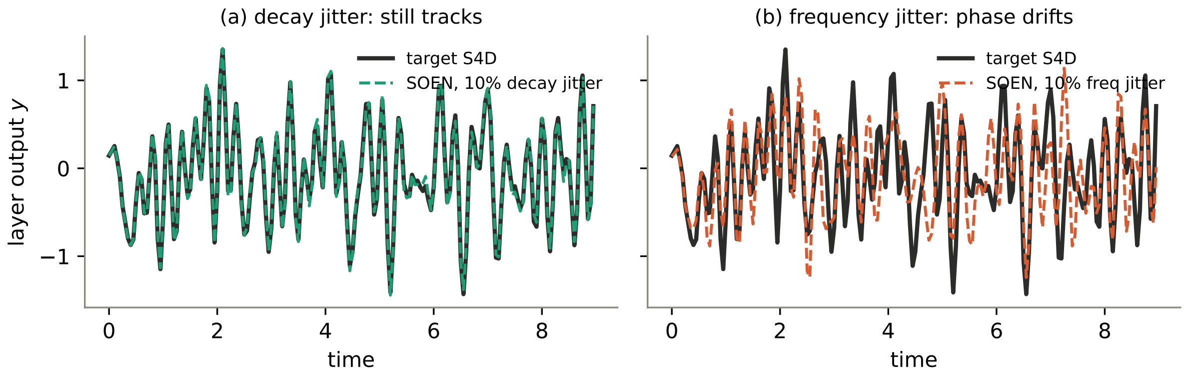

Figure 5: Example layer outputs (target S4D in black, SOEN realization dashed). (a) With 10% decay-timescale jitter the realization still tracks the target. (b) With 10% frequency jitter the envelope is preserved but the phase drifts—the visual signature of the frequency-precision constraint.

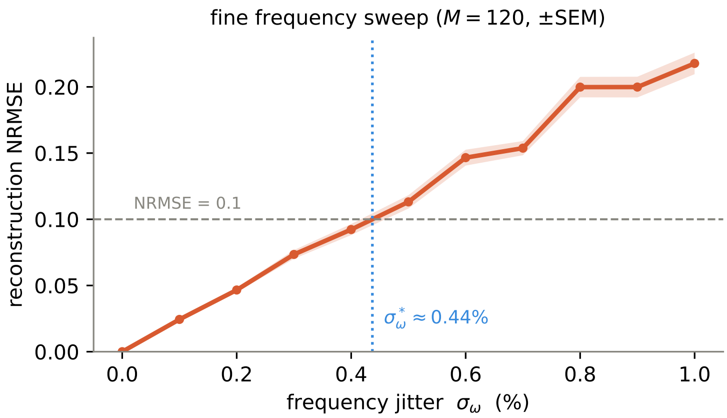

Figure 6: Fine frequency-jitter sweep ($M = 120$ device realizations per point, band is ±SEM). The reconstruction error crosses NRMSE $= 0.1$ at $\sigma_\omega^* \approx 0.45\%$, firming up the binding-constraint estimate from the coarse sweep.

to a resonator linewidth). The feature extractor is a bank of $M = 8$ sharp ($Q \sim 50$) SOEN resonator loops spanning the band; features are the log power per mode; a multinomial-logistic read-out is trained on them. *Naive* realization applies the ideally-trained read-out to jittered devices. *Device-aware* training instead fits the read-out to features pooled over many jittered device instances (jitter augmentation), with no change to the hardware.

Results (Fig. 7). The ideal read-out reaches 100% test accuracy. Under naive realization, accuracy falls with frequency jitter — 93%, 78%, 68% at $\sigma_\omega = 2, 3, 4\%$ — while decay jitter is benign out to 30%, mirroring the kernel result at the task level. Device-aware training recovers a large fraction of the loss: 68%→82% at $\sigma_\omega = 4\%$ and 61%→76% at 6%, by learning read-out weights that lean on robust ensembles of modes rather than knife-edge resonators.

Two messages. First, the frequency-precision constraint *persists* at the task level but is *softer* than the kernel metric: a magnitude (power) read-out over a redundant bank tolerates a few percent jitter, versus $\sim 0.45\%$ for phase-sensitive waveform reconstruction. Phase-insensitive read-outs and redundant banks buy robustness for free. Second, where naive realization does degrade, *device-aware training recovers much of it without touching the hardware* — the practical lever is to co-design the read-out with the device's measured jitter statistics.

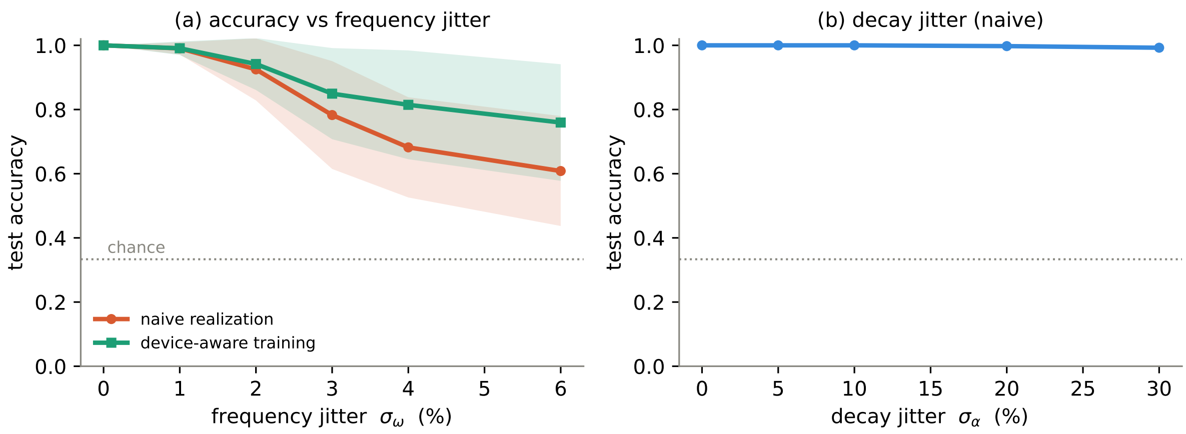

Figure 7: Task-level realization (3-class fine frequency discrimination; ideal accuracy 100%). (a) Naive realization (coral) loses accuracy as frequency jitter grows; device-aware training (teal), fitting the read-out to jitter-augmented features, recovers a large fraction. (b) Decay jitter is benign to 30% — frequency, not timescale, is the binding constraint at the task level too. Bands are ± 1 s.d. over device realizations; dotted line is chance.

9 Limitations and honest scope

This is a conceptual note, and the correspondence is only as good as its assumptions. (1) The exact mapping holds in the *linearized* regime (4); SOEN elements are deliberately nonlinear, and large-signal behavior departs from any fixed LTI mode. (2) Real devices have noise, parameter variability, and limited dynamic range; §7 quantifies the first two for a synthetic target, but dynamic range, quantization, and correlated (rather than independent) device errors remain unmodeled, and mode timescales are set by fabrication rather than a continuously trainable parameter. (3) Mapping a *trained* SSM onto hardware requires writing learned poles/timescales into physical τ and bias values; §8 demonstrates device-aware training of a linear read-out (reservoir-style), but end-to-end hardware-aware training that backpropagates through the device (e.g. via surrogate gradients), and at larger than CPU/toy scale, remains future work. (4) The selectivity claim (§5) is an analogy pending quantitative comparison. (5) Energy figures are order-of-magnitude and exclude cooling amortization. None of these void the dictionary; they delimit where it is exact, where it is approximate, and what must be measured to promote analogy to engineering.

10 Why it matters

The practical payoff is a shared language. An ML engineer who knows S4D can now read a SOEN datasheet as a set of fixed mode poles; a device physicist can borrow SSM initialization (HiPPO, S4D-Lin) as principled targets for the distribution of dendritic timescales. The realization experiment of §7 is a first instance of this program; the natural next steps are to drive it with a *trained* S4D layer on a real long-range task (rather than a synthetic target), and to train hardware-aware SSMs whose mode frequencies are regularized toward a fabricable, trimmable range given the $\sim 1\%$ tolerance found here. The companion interactive models let a reader watch a single loop's exponential kernel, stack modes into an S4D bank, and toggle a resonant dendrite into a complex mode—turning the dictionary in Table 1 into something one can manipulate.

References

[1] A. Gu, K. Goel, C. Ré, *Efficiently Modeling Long Sequences with Structured State Spaces (S4)*, ICLR (2022); arXiv:2111.00396.

[2] A. Gu *et al.*, *On the Parameterization and Initialization of Diagonal State Space Models (S4D)*, NeurIPS (2022); arXiv:2206.11893.

[3] A. Gu, T. Dao, *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*, arXiv:2312.00752 (2023).

[4] A. Orvieto *et al.*, *Resurrecting Recurrent Neural Networks for Long Sequences (LRU)*, ICML (2023); arXiv:2303.06349.

[5] A. Gu *et al.*, *HiPPO: Recurrent Memory with Optimal Polynomial Projections*, NeurIPS (2020); arXiv:2008.07669.

[6] J. M. Shainline *et al.*, *Superconducting optoelectronic loop neurons*, J. Appl. Phys. **126**, 044902 (2019).

[7] J. M. Shainline, B. A. Primavera, R. O'Loughlin, *Relating Superconducting Optoelectronic Networks to Classical Neurodynamics*, arXiv:2409.18016 (2024), NIST.

[8] A. Gu *et al.*, *Combining Recurrent, Convolutional, and Continuous-time Models with Linear State-Space Layers (LSSL)*, NeurIPS (2021); arXiv:2110.13985.

[9] S. M. Meyer *et al.*, *A Diagonal Structured State Space Model on Loihi 2 for Efficient Streaming Sequence Processing*, NeurIPS (2024); arXiv:2409.15022.

[10] S. Siegel, M.-J. Yang, J.-P. Strachan, *IMSSA: Deploying Modern State-Space Models on Memristive In-Memory Compute Hardware*, arXiv:2412.20215 (2024).

[11] X. Zhang, M. Hu, S. Lu, S. Kim, E. Y.-J. Lee, Y. Liu, W. D. Lu, *Compute-in-memory implementation of state space models for event sequence processing*, Nature Communications **17**, 1513 (2026).

[12] M. Bittner, A. Jantsch *et al.*, *Modeling Diagonal State Space Models as Electric Circuits for Analog Neural-Network Inference*, TU Wien (2025).