

Phase Structure in Continuous Wave Fields Enables Speech Classification Without Backpropagation

Jeremy Slater

dr.slater@stratusneuro.com

Stratus <https://orcid.org/0000-0002-2656-1354>

Gardar Thorvardsson

Kvikna Medical ehf

Article

Keywords: physical neural networks, Landau-Ginzburg dynamics, equilibrium propagation, wave computing, phase readout, keyword spotting, mel spectrogram

Posted Date: March 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9205518/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Yes there is potential Competing Interest. J.D.S. is Chief Medical Officer of Stratus Neuro, the Chief Innovation Officer for MERLN LLC, and founder of Medscrios LLC. G.T. is the managing director of Kvikna Medical.

Phase Structure in Continuous Wave Fields Enables Speech Classification Without Backpropagation

Jeremy D. Slater¹

Gardar Thorvardsson²

¹ MERLN LLC, Irving, TX, USA

dr.slater@stratusneuro.com

² Kvikna Medical, Reykjavik, Iceland

gardar@kvikna.com

Correspondence: dr.slater@stratusneuro.com

Abstract

Physical neural networks promise energy-efficient computation by exploiting the intrinsic dynamics of physical substrates, but existing implementations rely on discrete elements or restrict learning to digital readout layers. Here we demonstrate that a continuous Landau-Ginzburg wave field, operating in the underdamped regime, generates phase structure sufficient for classifying spoken words at 74.1% accuracy on a 10-class spoken command recognition task (Google Speech Commands V2) — using only a linear readout and no backpropagation through the physical dynamics. Systematic ablation across ten conditions reveals a three-tier hierarchy of contributions. First, operating in the underdamped regime accounts for ~20 percentage points (pp): a from-scratch baseline trained with EP but initialized in the overdamped regime ($\gamma/\omega = 1.0$) reaches only 53%, while theoretically-motivated underdamped initialization ($\gamma/\omega < 0.05$) yields 74.1% (EP fine-tuning contributed an additional 0.81 pp; ablation 7). Second, within the architecture, readout design is decisive: explicit phase extraction — cosine and sine of the phase angle, amplitude, and amplitude gradient — contributes 7.8 points over implicit complex-component encoding, revealing that linear readouts cannot exploit phase information unless it is explicitly projected into trigonometric form. Third, individual physics components — cross-phase modulation, spatial parameter grids, evanescent coupling, and EP fine-tuning of material parameters — each contribute less than 1 percentage point individually, indicating that the underdamped LG regime is robust to specific parameter choices once the architecture is correctly designed. Notably, Equilibrium Propagation drives the lateral inhibition strength toward opposite optima depending on the readout — higher under amplitude measurement ($D \rightarrow 0.27$), lower under phase-sensitive measurement ($D \rightarrow 0.025$) — demonstrating that EP co-adapts the

physical substrate to the measurement apparatus. Every learned parameter in principle maps directly to fabrication specifications for photonic or acoustic hardware.

Keywords: physical neural networks, Landau-Ginzburg dynamics, equilibrium propagation, wave computing, phase readout, keyword spotting, mel spectrogram

Introduction

The computational demands of modern deep learning have driven a search for alternatives to digital von Neumann architectures. Physical neural networks — systems that exploit the intrinsic dynamics of physical substrates to perform computation — offer a path toward substantially reduced energy consumption and latency by computing in the physics of wave propagation, diffusion, or mechanical resonance¹⁻³.

A central challenge in physical computing is the training problem: how does one adjust the physical substrate to improve performance? Backpropagation, the workhorse of digital deep learning, requires precise knowledge of gradients through every computational step — information that is generally unavailable in physical systems. Two dominant approaches have emerged in physical neural network demonstrations. The first fixes the physical substrate entirely and trains only a digital readout layer, treating the physics as a random or designed feature extractor⁴⁻⁷. The second employs perturbation-based methods that estimate gradients through finite differences, which require evaluations proportional to system dimensionality, though simultaneous perturbation approaches reduce this to two evaluations per step; however, the stochastic gradient estimates remain noisy at the parameter scales required for continuous field systems^{8,9}.

Equilibrium Propagation (EP) offers a third path. Originally proposed as a biologically plausible alternative to backpropagation for energy-based models¹⁰, EP computes parameter updates using only locally available information: the difference in a physical quantity between a free-running state and a weakly supervised state. Critically, EP requires no knowledge of the system's computational graph and operates through the same physics that performs the forward computation. Recent work has demonstrated EP in Ising machines¹¹ and coupled oscillator

networks¹², but these implementations use discrete elements that lack the rich spatiotemporal dynamics of continuous wave fields.

Here we ask: can a continuous wave field, governed by Landau-Ginzburg dynamics, learn to classify speech through Equilibrium Propagation? The Landau-Ginzburg (LG) equation describes a broad class of physical systems — from superconductors to nonlinear optical media to fluid convection — through a complex order parameter field whose dynamics include lateral inhibition, damping, intrinsic rotation, and nonlinear saturation. We show that this system, when operating in the underdamped regime (low damping-to-frequency ratio), generates spatiotemporal phase structure that a linear readout can decode for speech classification.

Our approach differs from prior physical neural networks in three respects. First, the computational substrate is a continuous complex-valued field rather than discrete coupled elements, providing access to the full richness of wave interference, lateral inhibition, and nonlinear phase dynamics. Second, EP adjusts spatially-resolved material parameters — lateral inhibition strength and damping rates defined on a spatial grid — that map directly to fabrication specifications for photonic or acoustic hardware. Third, we employ a phase-sensitive readout that extracts six features per spatial location, enabling the system to exploit the phase information that wave dynamics naturally generate.

We evaluate our system on Google Speech Commands V2¹³, a 10-class spoken command recognition task (see Methods). Systematic ablation across ten conditions reveals a three-tier hierarchy. The designed LG architecture is the dominant contribution: a from-scratch baseline trained with EP but without the designed field dynamics reaches only 53.0%, while the full system achieves 74.1% — a 21.1-point gap between the overdamped-initialization baseline and the full system, of which 20.3 points are attributable to underdamped initialization and 0.81 points to EP

fine-tuning. Within the architecture, readout design is the single most important choice: explicit phase extraction contributes 7.8 percentage points over implicit complex-component encoding, while amplitude-only measurement costs 4.1 points relative to the full phase-sensitive readout. Individual physics components — cross-phase modulation, spatial parameter grids, evanescent coupling, learnable rotation frequency, and EP fine-tuning of the 393 physics parameters — each contribute less than 1 percentage point, indicating that the underdamped LG regime is robust to specific parameter choices once the architecture is correctly designed.

Results

A Landau-Ginzburg wave field classifies speech at 74.1%

We constructed a three-channel LG wave field on a 94×64 spatial grid, driven by a mel spectrographic representation with delta and delta-delta temporal derivatives of one-second spoken word recordings (see Methods). Each channel operates at a distinct rotation frequency ($\omega \approx 0.9, 1.4, 2.1$ rad/step), spanning a logarithmic frequency hierarchy broadly analogous to the multi-scale oscillatory organization of cortical networks¹⁴⁻¹⁶. The field evolves for 120 timesteps per input according to the LG equation with spatially-resolved lateral inhibition strength $D(x,y)$, damping $\gamma(x,y)$, cubic saturation β , and cross-phase modulation (XPM) coupling between channels.

After settling, six features are extracted per spatial location per channel: the real and imaginary components of the field (Z_r, Z_i), the amplitude $|Z|$, the cosine and sine of the phase angle ($\cos \theta, \sin \theta$), and the local amplitude gradient R . These 27,072-dimensional feature vectors are classified by a delta-rule readout trained through local error correction requiring no backpropagation.

The full system achieves a peak validation accuracy of 74.1% and test accuracy of 73.22% on Google Speech Commands V2 with 10 command classes (known-class accuracy; silence and unknown excluded from evaluation, $N = 3,703$ validation samples; $N = 4,074$ test samples). Training proceeds in two phases: EP adjusts the 393 physics parameters over 20 epochs of joint physics-and-readout training, followed by readout optimization with frozen physics for 100 epochs. The best validation result occurs at readout epoch 21.

Phase-sensitive readout is the dominant design choice

To quantify the contribution of each architectural component, we performed systematic ablation experiments across ten conditions (Table 1, Fig. 1). Each ablation loads the trained checkpoint, removes or modifies one component, freezes all physics parameters, and retrain only the readout for 100 epochs.

The readout design is by far the most impactful factor, and the ordering of readout conditions reveals why. Implicit complex-component encoding (ablation 4: Z_r, Z_i only) reduces accuracy by 7.80 percentage points relative to the full six-feature readout, while amplitude-only measurement (ablation 9: $|Z|$ only) reduces accuracy by 4.10 points. The unexpected ordering — $|Z|$ alone outperforms Z_r, Z_i alone by 3.70 points — exposes a fundamental limitation of linear readouts: the phase relationship $\theta = \text{atan2}(Z_i, Z_r)$ is implicit in the complex components but inaccessible to a linear decoder. Z_r and Z_i act as correlated channels that partially cancel each other's discriminative contribution. Only explicit trigonometric projection ($\cos \theta, \sin \theta$) makes phase information linearly decodable. The amplitude gradient R provides additional spatial structure. Together, these explicit features account for 7.80 percentage points of the system's performance.

Individual physics components, by contrast, each contribute less than one percentage point. Removing cross-phase modulation (ablation 2) marginally improves accuracy by 0.08 points, indicating XPM does not contribute net discriminative benefit at this operating point. Truncating the rotation budget (ablation 6) costs 0.46 points, and removing evanescent coupling (ablation 3) costs 0.54 points — both within the noise band of the ablation methodology. Replacing spatially-resolved physics grids with scalar per-channel parameters (ablation 1) marginally improves accuracy by 0.43 points, indicating EP drove the lateral inhibition strength toward spatially

uniform values and spatial patterning of material parameters is not beneficial for this task. Fixing the rotation frequency (ablation 5) marginally improves accuracy by 0.22 points. Resetting all physics parameters to initialization without EP fine-tuning (ablation 8) costs 0.81 points — the largest individual physics contribution but still within one percentage point of the full system.

The picture that emerges is one of architectural robustness: once the LG field operates in the underdamped regime with phase-sensitive readout, the specific values of individual physics parameters matter little. The field generates useful phase structure across a broad range of parameter choices; extracting that structure explicitly is what determines performance.

The dynamical regime, not the learning rule, provides the dominant contribution

Comparing three key conditions reveals the relative contributions of architecture and learning (Fig. 2). Training from scratch with EP but initialized in the overdamped regime ($\gamma/\omega = 1.0$) achieves only 53.0% (ablation 7). The full LG architecture at theoretically-motivated initial parameters — before any EP fine-tuning — achieves 73.3% (ablation 8). The complete system with EP-optimized physics achieves 74.1% (ablation 0). The 21.1-point gap between baseline and full system decomposes as: 20.3 points from architecture (96%) and 0.81 points from EP fine-tuning (4%).

This result has a clear implication: the theoretical design choices — underdamped operation ($\gamma/\omega < 0.05$), multi-frequency rotation, and phase-sensitive readout — are collectively far more important than the learning algorithm. The LG dynamics at any reasonable parameterization in the underdamped regime generate a field rich in phase information. EP fine-tunes toward lower dissipation, but the architecture does the heavy lifting.

Equilibrium Propagation discovers a low-dissipation regime

Although EP's contribution is modest in magnitude and did not reach statistical significance (McNemar's test, $\chi^2 = 3.63$, $p = 0.057$, $n = 4,074$ test samples), it reveals a consistent and interpretable physical trajectory. Over 20 EP training epochs, the physics parameters evolve systematically toward lower dissipation.

Lateral inhibition strength decreases from $D = 0.100$ to $D = 0.025$ across all channels and spatial locations, approaching the lower bound. Reduced lateral inhibition preserves spatial structure in the field, preventing the suppression of phase patterns that carry classification information.

Damping decreases from $\gamma = [0.025, 0.050, 0.100]$ to $\gamma = [0.010, 0.013, 0.035]$, with the lowest-frequency channel reaching the imposed floor (0.010) by epoch 13. Lower damping extends the field's temporal memory: with $\gamma = 0.01$, the $1/\gamma$ time constant is approximately 100 timesteps, comparable to the full settling window. Rotation frequencies increase slightly from $\omega = [0.900, 1.365, 2.100]$ to $\omega = [0.912, 1.378, 2.124]$, and saturation coefficients decrease from $\beta = [0.060, 0.110, 0.210]$ to $\beta = [0.036, 0.086, 0.186]$.

The consistent direction of all parameter changes — toward lower dissipation, reduced lateral inhibition, higher rotation, lower saturation — reveals that EP discovers a single physical principle: minimize energy loss to preserve phase coherence. In the low-dissipation limit, the field approaches a conservative system where phase accumulates without degradation and spatial patterns persist without suppression.

The optimal physics depends on the measurement apparatus

A striking finding emerges from comparing the current results with earlier experiments using a scalar (amplitude-only) readout. With a scalar readout, EP drives lateral inhibition strength upward ($D \rightarrow 0.27$), blurring spatial structure to preserve only coarse features that the limited readout can exploit. With the phase-enhanced readout, EP drives lateral inhibition strength downward ($D \rightarrow 0.025$), preserving fine spatial structure because the readout can extract phase information from it.

Same field. Same task. Same learning rule. Different measurement apparatus yields opposite optimal physics. This demonstrates that EP co-adapts the material to the sensor, not just to the task. The physical substrate and the readout form a coupled system, and the optimal material specification depends fundamentally on what the readout can measure.

This has practical implications for photonic hardware design: the readout architecture must be specified before optimizing the physical substrate, because the optimal spatial coupling geometry depends on whether the detector measures amplitude, phase, or both.

Discussion

We have demonstrated that a continuous Landau-Ginzburg wave field achieves 74.1% accuracy on a 10-class speech classification task using zero backpropagation through the physical dynamics. The dominant contribution comes from the architecture itself — the combination of underdamped LG dynamics, multi-frequency rotation, and phase-sensitive readout generates a phase-rich field representation that a linear readout can decode.

Design principles for wave-based computing. Our ablation study yields three actionable principles for the physical computing community. First, operate in the underdamped regime ($\gamma/\omega < 0.05$): this maximizes the rotational character of the dynamics, generating rich phase trajectories that carry discriminative information. Second, measure phase explicitly: an amplitude-only readout costs 4.1 percentage points, and implicit complex-component encoding costs 7.8 points, because linear readouts cannot extract the phase relationship $\theta = \text{atan2}(Z_i, Z_r)$ without explicit trigonometric projection. Third, specify the readout before optimizing the physical substrate: EP co-adapts material parameters to the measurement apparatus, and the optimal lateral inhibition strength differs by an order of magnitude depending on whether the readout measures amplitude or phase.

Relation to neural oscillatory dynamics. The multi-frequency rotational dynamics in our system parallel observations in primate prefrontal cortex, where neural populations exhibit continuous rotational trajectories during cognitive tasks^{15,16}. The operating regime of our engineered system — underdamped, multi-frequency, phase-rich — converges on principles widely observed in cortical oscillatory dynamics, where frequency hierarchies spanning multiple scales support information integration across temporal and spatial domains¹⁴. The convergence

of an engineered wave computer and biological neural dynamics on similar operating regimes suggests that underdamped, phase-rich oscillatory systems may represent a universal computational motif independent of substrate.

Fabrication pathway. Every parameter in our system maps to a physical fabrication specification. The lateral inhibition strength D , implemented as a Difference-of-Gaussians spatial coupling kernel, corresponds to the spatial extent of inhibitory coupling in photonic lattices or acoustic metamaterials — physically realized through waveguide geometry and coupling gap specifications. The damping rate γ corresponds to material absorption or radiation loss. The rotation frequency ω corresponds to the carrier frequency or resonance frequency of the substrate. Our results specify the target regime: minimal lateral inhibition (small D), low loss (small γ), and multi-frequency operation spanning approximately one octave per channel. These specifications are within the range of current silicon photonics fabrication^{7,17}.

Limitations. The 74.1% accuracy, while notable for physical computing with local learning rules, remains below state-of-the-art deep learning systems on this benchmark (~97% with convolutional or attention-based architectures). This gap reflects both the constraint of zero backpropagation and the limited capacity of a single-layer linear readout operating on 393 learned physics parameters. The spatial physics grids marginally impaired performance when present (ablation 1, +0.43pp when replaced with scalar parameters), suggesting EP drove parameters toward spatially uniform values and that spatial material patterning is not beneficial for this task and readout combination. The overcomplete readout (27,072 dimensions for 11 classes) may compensate for suboptimal physics, potentially masking larger contributions from EP in more constrained settings. The unknown rejection mechanism achieved only 8.78%

detection rate and was not effective; a training strategy incorporating unknown samples or a contrastive objective on the phase coherence representation is a natural next step.

Future directions. A natural extension is multimodal integration, in which separate field layers process distinct sensory modalities — audio and vision — and a higher-order field learns to bind them through cross-modal ephaptic coupling. Such an architecture would test whether EP can discover inter-modal binding mechanisms that emerge from field dynamics rather than explicit design, and would provide a physically instantiated computational model of temporal binding via wave field resonance. Hardware implementation in photonic or acoustic substrates would validate the predicted performance and energy efficiency advantages of wave-based computing.

Methods

Dataset and evaluation protocol

We use Google Speech Commands V2, a dataset of one-second recordings sampled at 16 kHz. The command vocabulary comprises 10 words (yes, no, up, down, left, right, on, off, stop, go). The corpus also contains a silence class, sourced from random one-second crops of the six long ambient recordings in the `background_noise` directory, and an unknown category comprising all remaining vocabulary words not in the command set.

Because the `background_noise` source files lack the speaker-hash metadata used by the Google-standard train/validation/test assignment algorithm, all silence entries are assigned to the training split. Silence is accordingly present during training (as class index 10) but absent from validation and test sets, and silence classification is not evaluated. Training with silence prevents the field from developing spuriously high-confidence responses to ambient acoustic backgrounds; the class is masked to $-\infty$ at inference to prevent misclassification of command samples.

The readout maintains 11 output logits during training (10 command classes plus silence at index 10); at inference the silence logit is masked to $-\infty$ and is not evaluated. For all accuracy reporting the system is treated as a closed-set classifier over the 10 command classes only. Unknown samples are excluded from training entirely. The training set contains 30,775 samples across these 10 classes. Validation and test sets retain all 12 categories (9,981 and 11,005 samples respectively), with all non-command, non-silence samples pooled into a single unknown label; however, unknown samples are excluded from accuracy computation (see below).

Performance metrics. The primary reported metric — referred to throughout as known-class accuracy — is closed-set accuracy computed exclusively over the 3,703 validation samples (or 4,074 test samples) belonging to the 10 command classes. Unknown and silence samples are excluded from this computation. The 74.1% figure reported throughout this paper refers to this metric.

An adaptive coherence threshold was implemented for open-set unknown rejection. During training, the running mean and variance of the maximum logit score across the 10 class templates are tracked using Welford's online algorithm, and the rejection threshold is set to $\theta = \mu_{\text{coherence}} - 1.5\sigma_{\text{coherence}}$. Unknown detection rate reached 8.78% at the best checkpoint, indicating near-complete overlap between the score distributions of known commands and unknown sounds. Unknown rejection is therefore reported as a negative result and treated as an open direction for future work rather than a demonstrated capability of the current system.

Sensory front-end

Input audio is represented as a log-compressed mel spectrogram computed using a 400-sample Hann window with 160-sample hop length and 64 mel frequency bands, followed by per-sample z-score normalization. Three input channels are constructed from this representation: the log mel spectrogram (static spectral content), its first temporal derivative (onset and offset information), and its second temporal derivative (acceleration of spectral change). This standard three-channel representation provides the field with complementary temporal views of the acoustic signal across the 64 frequency bands and the temporal frames of each one-second recording. The 16

kHz audio with 400-sample window and 160-sample hop yields 101 time frames per one-second recording. The resulting $3 \times 101 \times 64$ representation (channels $\times$ time $\times$ frequency) is bilinearly interpolated to the 94×64 field resolution, with time mapped to the x-axis and frequency to the y-axis. The sensory drive $S_c(x,y,t)$ is applied as a constant additive input throughout all 120 settling steps — the field integrates over the full spectrogram simultaneously rather than receiving input frame-by-frame.

Landau-Ginzburg field dynamics

The complex-valued field $Z(x, y, t) \in \mathbb{C}$ evolves on a 94×64 spatial grid with three channels ($c = 0, 1, 2$) according to:

$$\partial Z_c / \partial t = -\gamma_c(x,y) Z_c + i\omega_{\text{eff},c} Z_c - \beta_c |Z_c|^2 Z_c + D_c(x,y) \cdot K_{\text{DoG}} * Z_c + \kappa_{\text{ev}} \sum_{\{c' \neq c\}} \exp(-|c-c'|) \cdot Z_{c'}(x,y,t) + S_c(x,y,t)$$

where $\omega_{\text{eff},c} = \omega_c + \chi_{\text{spm}} |Z_c|^2 + \chi_{\text{xpm}} \sum_{\{c' \neq c\}} |Z_{c'}|^2$ incorporates cross-phase modulation, $\gamma_c(x,y)$ is the spatially-resolved damping rate (8×8 grid, bilinearly interpolated to field resolution), ω_c is the rotation frequency per channel, β_c is the cubic saturation coefficient, and $S_c(x,y,t)$ is the sensory drive. The spatial coupling operator $K_{\text{DoG}} * Z_c$ is a Difference-of-Gaussians (Mexican Hat) lateral inhibition kernel¹⁸ (9×9 , $\sigma_{\text{exc}} = 1.0$, $\sigma_{\text{inh}} = 3.0$) with periodic boundary conditions, scaled by the spatially-resolved lateral inhibition strength $D_c(x,y)$ (8×8 grid, bilinearly interpolated), supporting spatial contrast enhancement through short-range excitation and longer-range inhibition. The evanescent coupling term $\kappa_{\text{ev}} \sum_{\{c' \neq c\}} \exp(-|c-c'|) \cdot Z_{c'}$ models near-field tunneling between adjacent-frequency waveguide

modes in photonic implementations; κ_{ev} is initialized to 0.05 and ablation 3 removes this term by setting $\kappa_{\text{ev}} = 0$. The field is integrated using semi-implicit Euler with implicit damping:

$Z_{\text{new}} = (Z + dt \cdot dZ) / (1 + dt \cdot \gamma)$, with $dt = 0.07$ for 120 steps.

Initial parameter values

Physics parameters are initialized from theoretical considerations of the underdamped regime.

Damping: $\gamma = [0.025, 0.050, 0.100]$ per channel. Rotation frequency: $\omega = [0.900, 1.365, 2.100]$

rad/step. Lateral inhibition strength: $D = 0.100$ (uniform). Saturation: $\beta = [0.060, 0.110, 0.210]$

(physical values after softplus transform with $|\text{raw}| + 0.01$ floor). XPM coefficients: $\chi_{\text{spm}} =$

0.269, $\chi_{\text{xpm}} = 0.158$ (physical values after sigmoid transform). Evanescent coupling: $\kappa_{\text{ev}} =$

0.05. These values place the system in the underdamped regime with γ/ω ratios between 0.03 and 0.05.

Equilibrium Propagation

EP proceeds in two phases per training sample. In the free phase, the field settles from the sensory drive for 120 timesteps without supervision. In the nudged phase, a weak supervisory signal ($\beta_{\text{nudge}} = 0.1$) back-projects the readout error through the weight matrix into field space, applied at every settling step after upsampling from the pooled 47×32 to the full 94×64 resolution. Physics parameters are updated proportionally to the difference in local field statistics between phases:

$$\Delta D(x,y) \propto [\text{Re}(K_{\text{DoG}} * Z \cdot Z^*)]_{\text{nudged}} - [\text{Re}(K_{\text{DoG}} * Z \cdot Z^*)]_{\text{free}}$$

$$\Delta \gamma(x,y) \propto [|Z|^2]_{\text{nudged}} - [|Z|^2]_{\text{free}}$$

$$\Delta \beta \propto [|Z|^4]_{\text{nudged}} - [|Z|^4]_{\text{free}}$$

$$\Delta \omega \propto [\text{Im}(Z^* \cdot dZ/dt)]_{\text{nudged}} - [\text{Im}(Z^* \cdot dZ/dt)]_{\text{free}}$$

All updates are local: each grid cell adjusts based only on the field values at that location.

Updates are clipped to ± 0.01 per step. No global error signal or gradient computation is required.

EP learning rate is 0.01 for physics parameters. Parameters are clamped to physically meaningful ranges ($D \in [0.001, 1.0]$, $\gamma \in [0.01, 0.5]$).

Phase-enhanced readout

After settling, the field is spatially pooled (2×2 average pooling, yielding 47×32 locations) and six features are extracted per location per channel: (1) Z_r : real part; (2) Z_i : imaginary part; (3) $|Z|$: amplitude; (4) $\cos \theta$: cosine of phase angle; (5) $\sin \theta$: sine of phase angle; (6) R : local amplitude gradient (Sobel filter magnitude). After 2×2 average pooling, the 94×64 field reduces to 47×32 spatial locations; with 3 channels and 6 features per location, this yields $47 \times 32 \times 3 \times 6 = 27,072$ dimensions total. The resulting 27,072-dimensional vector is classified by a delta-rule readout trained by minimizing the cross-entropy error signal: $\Delta W = \eta \times (\text{one_hot}(\text{label}) - \text{softmax}(Wf + b)) \otimes f_{\text{normalized}}$, where $f_{\text{normalized}}$ is the feature vector normalized by Welford's online statistics, with $\eta = 0.005$. Bias terms are updated correspondingly. At inference,

the silence class logit (index 10) is masked to $-\infty$ before argmax to prevent spurious assignment of command samples to the untested silence class.

Ablation methodology

Each ablation loads the best trained checkpoint (epoch 20 of joint EP+readout training, 70.73% validation accuracy), modifies one component, freezes all physics parameters, reinitializes the readout weights to zero, and retrain the readout for 100 epochs. The full system control (ablation 0) follows the same protocol without modification, establishing a 74.10% peak validation accuracy baseline that reflects the benefit of readout optimization against stable physics. The from-scratch baseline (ablation 7) trains both EP physics and readout from default initialization ($\gamma = \omega = 0.5$ per channel, $\gamma/\omega = 1.0$; overdamped regime) for 200 epochs, using the identical LG field architecture and EP update rules as the full system. The initial-physics ablation (ablation 8) resets all physics to initialization values while retaining the full architecture and phase readout. The statistical significance of the difference between ablation 0 and ablation 8 was evaluated using McNemar's test with continuity correction on test set predictions ($n = 4,074$ known-class test samples). The contingency table showed 2,522 samples correct under both conditions, 687 incorrect under both, 461 correct only under ablation 0, and 404 correct only under ablation 8 (865 discordant pairs total). This yields $\chi^2 = 3.63$, $p = 0.057$, indicating the EP contribution, while consistent in direction, does not reach significance at conventional thresholds.

Computational resources

All experiments were conducted on a single NVIDIA GPU (DGX Spark). Each ablation requires approximately 90 minutes for field precomputation (one-time forward pass through frozen physics for all training samples) followed by approximately 60 minutes for 100 epochs of readout training. The from-scratch baseline (ablation 7) requires approximately 20 hours for 200 epochs of joint EP and readout training. Total computational cost for all experiments: approximately 48 GPU-hours.

Data availability

Google Speech Commands V2 is publicly available at <https://arxiv.org/abs/1804.03209> (http://storage.googleapis.com/download.tensorflow.org/data/speech_commands_v0.02.tar.gz).

Trained checkpoints will be deposited in a public repository upon publication.

Code availability

Complete source code for the FICU architecture, training pipeline, and ablation framework will be made available in a public repository upon publication.

Acknowledgements

J.D.S. thanks Earl K. Miller, whose 2025 Society for Neuroscience Presidential Lecture, 'Cognition Emerges From Neural Dynamics,' prompted the line of inquiry that led to this work.

Author contributions

J.D.S. conceived the architecture, designed the experiments, performed all computational work, analyzed the data, and wrote the manuscript. G.T. provided mathematical review and critical analysis of the theoretical framework, verified the mathematical consistency of the field dynamics formalism, and contributed to manuscript revision through iterative discussion of the wave-interaction model of cognition.

Use of AI Statement

Artificial intelligence tools (Claude, Anthropic) were used to assist with computational simulation development and manuscript preparation. All AI-generated content was reviewed, validated, and revised by the author, who takes full responsibility for the accuracy and integrity of the work.

Competing interests

J.D.S. is Chief Medical Officer of Stratus Neuro, the Chief Innovation Officer for MERLN LLC, and founder of Medscrios LLC. G.T. is the managing director of Kvikna Medical.

Tables

Table 1 | Ablation study results. Each row shows the effect of removing or modifying one architectural component. Physics parameters are frozen at their trained values (except ablations 7 and 8), and only the readout is retrained. Δ indicates difference from the full system control. All accuracies are known-class accuracy computed over the 10 command classes only ($N = 3,703$ validation samples). Silence and unknown samples are excluded. Ablation 4 uses the real and imaginary field components only (Z_r, Z_i); ablation 9 uses amplitude only ($|Z|$). The ordering ablation 9 > ablation 4 demonstrates that implicit phase encoding is less accessible to linear readouts than amplitude alone.

#	Configuration	Peak val. acc. (%)	Δ (pp)	Test acc. (%)
0	Full system (control)	74.10	—	73.22
1	Spatial $\rightarrow$ scalar physics	74.53	+0.43	—
2	No cross-phase modulation	74.18	+0.08	—
3	No evanescent coupling	73.56	-0.54	—
4	Complex components only (Z_r, Z_i)	66.30	-7.80	—
5	Fixed rotation frequency	74.32	+0.22	—
6	Truncated rotation (259°)	73.64	-0.46	—
7	From-scratch baseline	53.04	-21.06	—
8	Initial physics (no EP)	73.29	-0.81	—
9	Amplitude-only readout ($ Z $)	70.00	-4.10	—

Figure Legends

Fig. 1 | Component ablation study sorted by impact. Bars show peak validation accuracy for the full system (blue), each ablation condition (red), EP physics reset (orange), and from-scratch baseline (grey). Numbers above bars indicate delta from full system control; green values indicate conditions where removal improves performance. The readout design dominates: explicit phase extraction contributes 7.80 pp over implicit complex-component encoding (ablation 4), while amplitude-only measurement costs 4.10 pp (ablation 9). Individual physics components — XPM, evanescent coupling, spatial grids, rotation frequency — each contribute less than 1 pp. Resetting all physics to initialization (orange) costs only 0.81 pp, indicating the underdamped LG regime is largely accessible without EP fine-tuning.

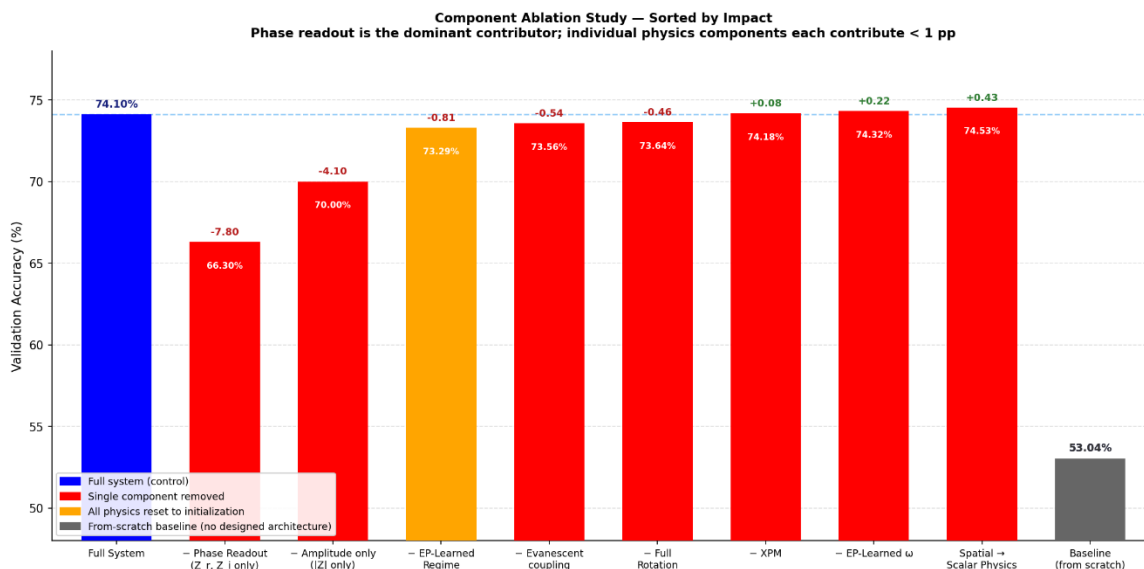


Fig. 2 | Architecture provides the dominant contribution; EP fine-tunes. Three-bar decomposition of performance gains. The from-scratch baseline (grey, 53.04%) represents EP training without the designed LG architecture. The LG architecture at theoretically-motivated initial parameters (orange, 73.29%) accounts for 20.25 of the 21.06 percentage point total improvement (96%). EP fine-tuning of the physics (blue, 74.10%) provides an additional 0.81 points (4%), a difference that did not reach statistical significance (McNemar's test, $p = 0.057$). The phase-sensitive readout alone accounts for 7.80 of the 21.06 total points (37%).

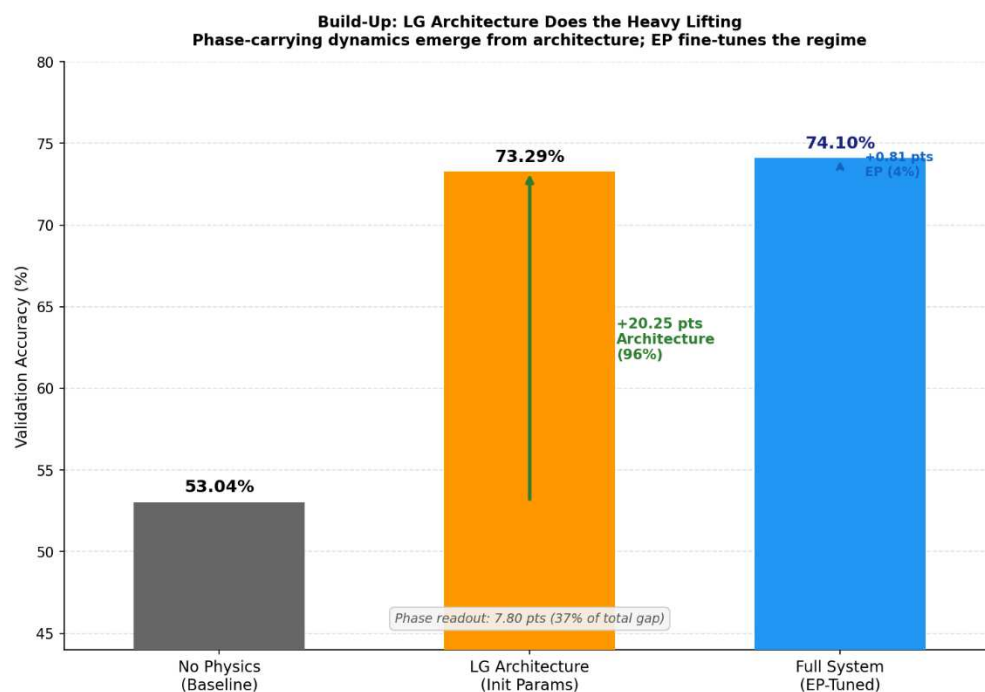
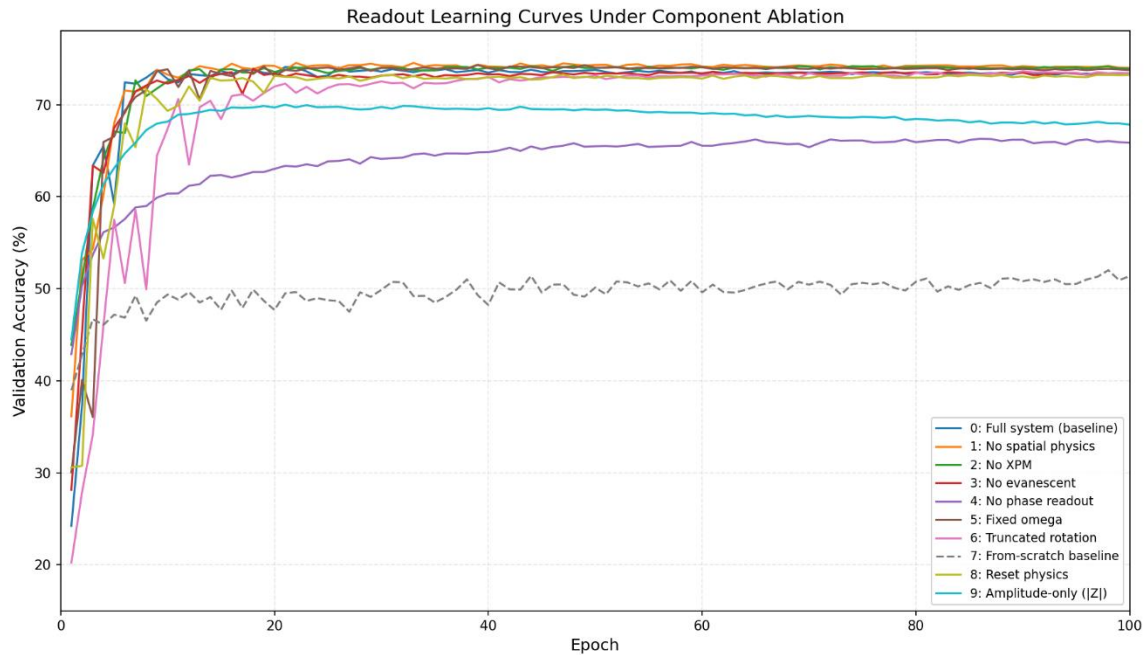


Fig. 3 | Readout learning curves under component ablation. Validation accuracy over 100 epochs of readout training with frozen physics. Three distinct performance tiers are visible. The main cluster (ablations 0–3, 5–6, 8) converges tightly between 73–75%, demonstrating that the LG dynamics generate a robust representation across a wide range of parameter and component variations. The amplitude-only condition (ablation 9, cyan, $|Z|$) plateaus at $\sim 70\%$, 4.1 points below the full system — confirming that amplitude alone carries substantial but incomplete discriminative information. The implicit complex-component condition (ablation 4, purple, Z_r and Z_i only) plateaus at $\sim 66\%$, 3.7 points below amplitude-only, demonstrating that phase information encoded implicitly in complex components is less accessible to a linear readout than explicit amplitude measurement. The from-scratch baseline (ablation 7, dashed grey) plateaus at $\sim 53\%$, establishing the floor attributable to underdamped initialization. Most conditions reach asymptotic performance within 20 epochs.



REFERENCES

- 1 Momeni, A., Rahmani, B., Mallejac, M., Del Hougne, P. & Fleury, R. Backpropagation-free training of deep physical neural networks. *Science (New York, N.Y.)* **382**, 1297–1303 (2023). <https://doi.org/10.1126/science.adi8474>
- 2 Wright, L. G. *et al.* Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022). <https://doi.org/10.1038/s41586-021-04223-6>
- 3 Stern, M., Arinze, C., Perez, L., Palmer, S. E. & Murugan, A. Supervised learning through physical changes in a mechanical system. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 14843–14850 (2020). <https://doi.org/10.1073/pnas.2000807117>
- 4 Tanaka, G. *et al.* Recent advances in physical reservoir computing: A review. *Neural Netw* **115**, 100–123 (2019). <https://doi.org/10.1016/j.neunet.2019.03.005>
- 5 Nakajima, K. Physical reservoir computing—an introductory perspective. *Japanese Journal of Applied Physics* **59**, 060501 (2020). <https://doi.org/10.35848/1347-4065/ab8d4f>
- 6 Larger, L. *et al.* Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing. *Optics express* **20**, 3241–3249 (2012). <https://doi.org/10.1364/OE.20.003241>
- 7 Vandoorne, K. *et al.* Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat Commun* **5**, 3541 (2014). <https://doi.org/10.1038/ncomms4541>
- 8 Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control* **37**, 332–341 (1992). <https://doi.org/10.1109/9.119632>
- 9 Bandyopadhyay S, Sludds A, Krastanov S, Hamerly R, Harris N, Bunandar D, Streshinsky M, Hochberg M, Englund D. Single chip photonic deep neural network with accelerated training. *arXiv>Emerging Technologies>Optics* (2022). <https://doi.org/10.48550/arXiv.2208.01623>
- 10 Scellier, B. & Bengio, Y. Equilibrium Propagation: Bridging the Gap between Energy-Based Models and Backpropagation. *Front Comput Neurosci* **11**, 24 (2017). <https://doi.org/10.3389/fncom.2017.00024>
- 11 Laydevant, J., Markovic, D. & Grollier, J. Training an Ising machine with equilibrium propagation. *Nat Commun* **15**, 3671 (2024). <https://doi.org/10.1038/s41467-024-46879-4>
- 12 Wang, Q., Wanjura, C. C. & Marquardt, F. Training coupled phase oscillators as a neuromorphic platform using equilibrium propagation. *Neuromorphic Computing and Engineering* **4**, 034014 (2024). <https://doi.org/10.1088/2634-4386/ad752b>
- 13 Warden, P. Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition. *arXiv>Computer Science>Computation and Language* (2018). <https://doi.org/10.48550/arXiv.1804.03209>
- 14 Buzsáki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science (New York, N.Y.)* **304**, 1926–1929 (2004). <https://doi.org/10.1126/science.1099745>
- 15 Aoi, M. C., Mante, V. & Pillow, J. W. Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nature neuroscience* **23**, 1410–1420 (2020). <https://doi.org/10.1038/s41593-020-0696-5>

- 16 Libby, A. & Buschman, T. J. Rotational dynamics reduce interference between sensory and memory representations. *Nature neuroscience* **24**, 715–726 (2021).
<https://doi.org/10.1038/s41593-021-00821-9>
- 17 Bogaerts, W. *et al.* Programmable photonic circuits. *Nature* **586**, 207–216 (2020).
<https://doi.org/10.1038/s41586-020-2764-0>
- 18 Wilson, H. R. & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal* **12**, 1–24 (1972).
[https://doi.org/10.1016/S0006-3495\(72\)86068-5](https://doi.org/10.1016/S0006-3495(72)86068-5)