A CROSSOVER, NOT A CONSTANT: WHEN COMPLEX-VALUED NETWORKS ACTUALLY HELP?

Anonymous authors
Paper under double-blind review

ABSTRACT

Complex-valued neural networks (CVNNs) are widely motivated in domains where the information is carried in magnitude and phase. The paper asks when that motivation is borne out and finds that the answer is a narrow regime and not a general advantage. On a geometry-controlled seven-class RadioML 2018.01A task, covering phase-dominant, amplitude-dominant, and joint modulations across nine SNR levels, a complex model beats the best real baseline by 7.94 pp against the Cartesian-only baseline set conventionally used in this literature, and by 0.71 pp once a polar real baseline is included, then *loses* to it by 0.58 pp under independent per-family tuning of the same search space, and by 3.29 pp under a stable complex activation. What is left is not a fixed advantage but a *crossover*: complex arithmetic is advantageous up to 12 pp in the marginal-SNR regime where phase estimates are noisy and is overtaken by an explicit polar representation as soon as the signal is clean. For all three configurations we test, the crossover is between 2 and 6 dB, regardless of the class composition and the sequence length. Three separate confounds inflate the apparent advantage outside this regime: (i) an impoverished baseline set, under which the capacity-matched Cartesian baselines become seed-unstable at the selected learning rate while polar and magnitude baselines do not; (ii) the selection rule, since matched-shared-trial selection measures robustness to a shared trial rather than peak performance; and (iii) training budget, since a 9.3 pp complex advantage on phase-shift-keying data falls to 0.4 pp at convergence. A fourth axis is not a confound but a design choice with confound-sized effects: complex activation moves accuracy by up to 61 pp on amplitude-structured EEG tasks, and the sign of the effect is task-dependent. Especially, $|z|$-gated activations recover amplitude structure but degrade phase-dominant tasks across RF, quantum, and EEG data, so magnitude-gating is a diagnostic to report rather than a prescription to follow. Finally, we reinterpret the complex layer architecture as a $U(1)$-equivariant subspace of a stacked-real two-channel layer architecture, not a gain but a reduction in expressiveness in line with symmetry, and test this interpretation by constructing a constrained real network and finding that it matches the complex network to a tolerance of -0.9 pp while outperforming a matched unconstrained real network by 2.4 to 11.1 pp. In attributing results to complex-valued arithmetic, it is the symmetry constraint that matters, not the data type. The code and results are available here [1].

1 INTRODUCTION

Many domains natively produce complex data $z = x + iy = re^{i\theta}$, including RF quadrature samples, Fourier transforms, quantum wavefunctions, and analytic neural time series. Consequently, Complex-Valued Neural Networks (CVNNs) are an intuitively appealing choice, often matching or outperforming real-valued models in signal classification Hirose (2006); Trabelsi et al. (2017), MRI Virtue et al. (2017); Cole et al. (2021), and unitary recurrent modeling Arjovsky et al. (2016); Wisdom et al. (2016).

However, complex inputs alone do not explain why a CVNN should intrinsically perform better. Real-valued networks can process the exact same information via standard Cartesian (x, y) or ex-

[1] https://anonymous.4open.science/r/Neural-Networks-in-Complex-Spaces-C1C2/

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

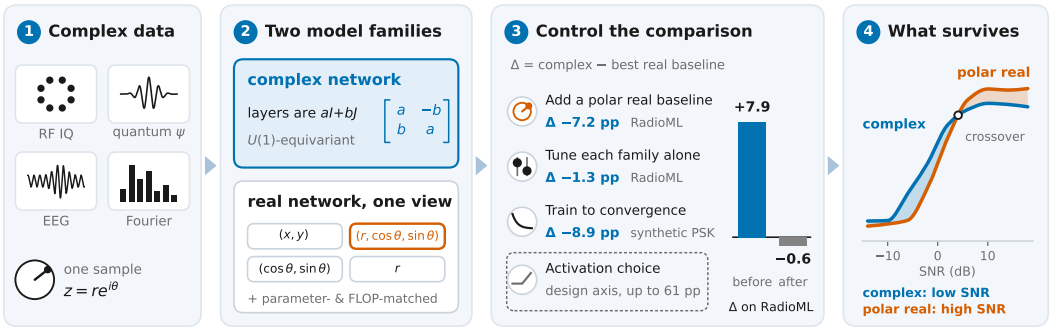


Figure 1: **Overview.** (1) A complex sample $z = re^{i\theta}$ is given (2) to a complex network ($U(1)$-equivariant $aI{+}bJ$ layers) and to real networks that each see one coordinate view. (3) Controlling the comparison takes the RadioML gap Δ from $+7.9$ to -0.6 pp; each control is shown with its measured effect and task. (4) What survives is an SNR crossover between 2 and 6 dB.

plicitly decoupled polar $(r, \cos\theta, \sin\theta)$ coordinates. Therefore, any measured CVNN advantage might not stem from complex arithmetic itself. It could instead result from a more favorable coordinate view, an implicit symmetry constraint, or mere robustness to suboptimal hyperparameter regimes that destabilize real-valued baselines.

This work poses a more precise question than the standard "Are complex-valued networks superior to real-valued ones?" This paper asks: *in which specific regimes do they offer a tangible advantage, what is the magnitude of the benefit, and what mechanisms explain the residual performance gap?* The answer consists of a positive and a negative half. The positive half identifies a distinct signal-to-noise ratio (SNR) crossover (Section 4). The negative half demonstrates that, outside of this specific regime, the perceived CVNN advantage is driven entirely by three methodological confounds that are individually understood but rarely isolated within this literature. The core contributions are:

- **An SNR-dependent crossover**: On RadioML O'Shea et al. (2018) data, the CVNN advantage peaks at $+12$ pp near -6 dB and subsequently reverses between 2 and 6 dB across all tested configurations.
- **Isolation of three distinct confounds**: Baseline coordinate systems, selection rules, and training budgets independently inflate the measured performance gap. Removing them shifts the gap to -0.58 pp.
- **Attribution to symmetry**: A $U(1)$-constrained real-valued network matches the complex model's performance within 0.9 pp and substantially outperforms a parameter-matched unconstrained real-valued model by $2.4 - 11.1$ pp, proving the symmetry constraint, not complex arithmetic, drives performance.

2 Background

Multiplying $x + iy$ by $a + ib$ operates on stacked real coordinates as the transformation matix $aI + bJ$ where $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. This forms a two-parameter subspace constrained within the broader four-parameter space of general real 2×2 channel mixings.

Proposition 1 (Equivariant two-channel kernels are complex scalars.)**.** *For a weight matrix $W \in \mathbb{R}^{2\times 2}$, the following properties are mathematically equivalent: (1) $WR_\phi = R_\phi W$ for all rotation angles ϕ; (ii) W can be expressed as $aI{+}bJ$; (iii) W represents scalar multiplication by the complex number $a + ib$.*

This represents the standard characterization of the commutant of $U(1)$ and is implicit within the steerable-CNN literature Cohen & Welling (2016); Worrall et al. (2017); we restate it here to formalize the interpretation utilized throughout our analysis. A complex `Conv1d` layer is intrinsically *not* more expressive than a stacked-real two-channel layer. It is, in fact, *less* expressive in a highly structured manner: it explicitly removes the degrees of freedom that violate global phase equivariance. A constraint of this nature proves highly beneficial when the unconstrained degrees of freedom would

otherwise overfit to noise (precisely the characteristic of the marginal-SNR regime). Conversely, it acts as a bottleneck once the data is sufficiently clean to accurately estimate those extra parameters. This dynamic perfectly maps to the crossover measured in Section 4. Furthermore, because this constraint can be wholly expressed using standard real arithmetic, it is completely separable from the complex datatype itself. Section 4.4 rigorously tests this separation, and formal proofs are provided in Appendix A.

2.1 The Liouville Activation Trilemma

Proposition 2 (Activation trilemma)**.** *No activation function $\sigma : \mathbb{C} \to \mathbb{C}$ can be simultaneously bounded on all of $\mathbb{C}$, entire (holomorphic over the whole plane) and non-constant.*

This follows immediately from Liouville's theorem. Every functional non-constant complex activation must therefore sacrifice either boundedness or holomorphy. The mathematical taxonomy this constraint induces is not merely theoretical; it directly categorizes activations based on *how* they interact with the magnitude $|z|$. As Section 5 demonstrated, this axis carries immense first-order empirical weight. We emphasize that this taxonomy successfully predicts *that* the activation choice will matter significantly, rather than strictly dictating which choice is universally best; the sign of the performance effect is heavily dependent on whether the ground truth relies primarily on amplitude or phase.

Holomorphy is not a prerequisite for network training. A complex activation function does not need to be holomorphic to be successfully optimized. Modern autodiff frameworks treat $f : \mathbb{C} \to \mathbb{C}$ as a standard real-differentiable mapping of $(\Re z, \Im z)$. For a real-valued loss function, gradient descent computes the conjugate Wirtinger derivative of $\partial f / \partial \bar{z}$, which remains well-defined regardless of whether the Cauchy-Riemann condition $(\partial f / \partial \bar{z} = 0)$ holds. Therefore, every activation we evaluate is mathematically admissible for backpropagation, including strictly magnitude-gated functions that are nowhere holomorphic. Proposition 2 dictates what a complex activation can mathematically *be*, not whether it can be computationally trained. Full details are provided in Appendix B.

3 Evaluation Protocol

Rather than declaring complex networks generally "win" or "lose," our protocol isolates the exact mechanisms driving measured performance gaps across five diagnostic axes: **Representation** (ruling out coordinate-view artifacts via standard Cartesian (x_t, y_t), polar $(|z_t|, \cos\theta_t, \sin\theta_t)$, phase-only, and magnitude-only baselines); **Symmetry** (isolating the $U(1)$ constraint from the datatype, see Section 4.4); **Capacity** (ruling out parameter/FLOP advantages via mathematically matched real baselines); **Selection rule** (separating hyperparameter robustness from true peak performance); and **Invariance** (testing end-to-end rotation survival). Note that we utilize $(\cos\theta, \sin\theta)$ rather than raw θ to bypass the branch cut discontinuity.

Selection rules and their application. We explicitly contrast two evaluation methodologies that must never be conflated. *Matched-shared-trial* selection evaluates all baselines at the complex model's single best validation trial, measuring robustness under a shared hyperparameter allocation. *Independent per-family* selection allows each model family to independently select its optimal trial, estimating true peak performance. When results from these rules diverge, the divergence itself is our primary finding. For RadioML (Section 4), we draw 16 random trials per family and report both metrics. Conversely, the pilots in Section 5 deliberately employ a static, shared learning rate (0.01) to explicitly demonstrate the exact benchmarking flaw this paper criticizes. We measured this confound directly: across a massive $100\times$ learning rate sweep, no setting ever lifts the complex model above 0.388 on amplitude events or 0.340 on coupling, while the explicit polar baseline effortlessly reaches 0.958 (Table 10). Consequently, no hyperparameter setting closes the fundamental gap reported in that section.

4 RadioML: A Crossover and Three Confounds

Experimental Setup. We utilize a highly controlled subset of the RadioML 2018.01A dataset O'Shea et al. (2018) spanning three distinct modulation geometries: phase-dominant (BPSK,

Table 1: Complex minus the best-real performance gap on the geometry-controlled RadioML subset (CReLU activation unless noted). All results share the exact same underlying data, the same 16-trial search space, and the same five random seeds; the only variable changing is the evaluation protocol.

Protocol	Best real baseline	Gap (pp)
Matched-shared-trial, Cartesian-only baseline set	FLOP-matched, 0.358	$+7.94$
Matched-shared-trial, all seven families	polar, 0.430	$+0.71$
Independent per-family, all seven families	polar, 0.443	-0.58
Independent per-family, ZReLU	polar, 0.443	-3.29

QPSK, 8PSK), amplitude-dominant (4ASK), and joint modulations (16QAM, 64QAM, 16APSK). We test across nine SNR levels from -14 to $+18$ dB, allocating 256 examples per class per SNR, with a standard sequence length of 128. We comprehensively sweep 16 random trials $\times$ 5 seeds $\times$ 7 distinct model families and report outcomes under both selection rules.

Rigorous Class-Order Verification. The class list originally distributed with the RadioML archive is documented to contain an incorrect one-hot ordering; utilizing it blindly would mislabel every downstream experiment without triggering a software error. We explicitly verified the index-to-modulation mapping against objective signal statistics at 30 dB. The underlying amplitude structure cleanly separates the classes: the coefficient of variation (CV) for $|z|$ falls monotonically across OOK/4ASK/8ASK/BPSK (0.767/0.634/0.579/0.462). OOK possesses the highest near-zero sample fraction (11.1% of samples fall below a tenth of the per-example mean amplitude). FM and GSMK correctly register as constant-envelope signals (CV=0.005 and 0.085). Furthermore, an M-th power test definitively confirms BPSK at order 2 and QPSK at order 4. However, this test does *not* reliably resolve 8PSK and higher orders: at 1024 samples under this specific archive's carrier-frequency offset, every PSK order exhibits the exact same weak second-order line and lacks a decisive high-order line. Therefore, we do not rely on it beyond QPSK. The fully verified 24-class table is provided in Appendix D, and the corrected order is strictly pinned so no experimental run relies on the faulty default loader.

4.1 The Measured Gap is Dictated by the Controls

Table 1 systematically separates the experimental confounds. Of the initial 7.94 pp advantage measured against the limited Cartesian-only baseline set, 7.22 pp can be attributed directly to the restrictive baseline set itself. A further 1.29 pp is an artifact of the selection rule. This leaves a true residual of -0.58 pp: once rigorous experimental controls are applied, the complex model actually falls *behind* the best real-valued baseline. The matched shared trial inherently selects a specific learning rate of 0.0236. At this specific rate, the capacity-matched Cartesian baselines become highly unstable across random seeds (showing standard deviations of 0.045 and 0.053), whereas the polar and magnitude real baselines remain perfectly stable (0.021 and 0.013). **The explicit polar real representation is fundamentally robust to the exact hyperparameter regime that severely destabilizes Cartesian baselines**. While stacked real remains stable on this specific task (0.018), it is simply weaker overall (0.336). The catastrophic seed collapse driving the mechanism described in Section 4.3 is the most prominent on the three-class subset, where the stacked real baseline exhibits a standard deviation of 0.134.

4.2 The Residual Effect: A Signal-to-Noise Crossover

Three consistent properties hold across every evaluated configuration (Table 2). First, the crossover point consistently sits between 2 and 6 dB regardless of the underlying class composition or the input sequence length. Second, the absolute peak of the complex advantage always manifests at -6 or -2 dB. Third, under rigorous independent per-family selection, the complex model lags behind the best real baseline in *every single* configuration tested.

Two specific experimental controls further sharpen this interpretation. First, introducing three amplitude-heavy classes (32QAM, 128QAM, 256QAM) predictably reduces the peak complex ad-

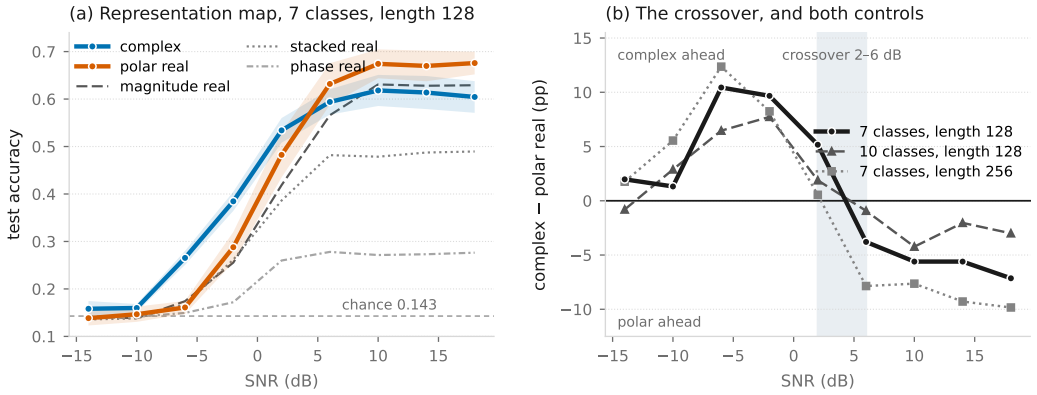


Figure 2: **The complex advantage is an SNR-dependent crossover, not a constant structural superiority.** *(a)* Per-SNR test accuracy parsed by representation on the geometry-controlled seven-class subset, utilizing matched-shared-trial selection and CReLU activation, averaged over five seeds. Shaded bands represent ± 1 standard deviation across seeds for the specific two families defining the crossover. *(b)* The explicit performance delta (Complex minus polar real), measured in percentage points, across configurations. The crossover manifests as a clear zero crossing. Every tested configuration invariably changes sign inside the highlighted 2 to 6 dB band, remaining entirely agnostic to class composition or sequence length.

Table 2: Complex model accuracy minus polar real baseline accuracy, evaluated per SNR, across three strictly independent test configurations. The crossover band is visually shaded in Figure 2.

Configuration	Peak adv. (pp)	At 18 dB (pp)	Crossover	Independent gap (pp)
7 classes, length 128	$+10.4$	-7.1	2-6 dB	-0.58
10 classes, length 128	$+7.7$	-3.0	2-6 dB	-0.60
7 classes, length 256	$+12.4$	-9.8	2-6 dB	-0.93

vantage from +10.4 to +7.7 pp, as the phase-dominant fraction of the overall task decreases. This is exactly the behavior the representation framework predicts and the data confirms. Second, doubling the input sequence length symmetrically amplifies *both* sides of the crossover (+12.4 pp at low SNR, -9.8 pp at high SNR). More temporal observation aids the complex model when the phase is heavily obscured by noise, but it aids the explicit polar model even more once the phase signal is pristine.

4.3 The Mechanism is Learning Rate, Not Gradient Magnitude

Prior literature frequently attributes real-baseline instability under matched-shared-trial selection to excessively large first-step gradients. However, under Adam and AdamW at $t = 1$, the bias-corrected moments yield an exact initial update of $\Delta w = -\eta \operatorname{sign}(g)$. Thus, the first step shifts every coordinate by exactly η, strictly independent of the actual gradient magnitude. We verified this numerically: scaling the loss landscape by $100\times$ or $0.01\times$ leaves $|\Delta w| = \eta$ perfectly unchanged down to 10^{-6} precision.

Consequently, gradient clipping cannot rescue a first-step failure. Empirically, clipping the global gradient norm to 1.0 on the three-class configuration leaves Cartesian instability entirely intact (seed standard deviations remain high at ~ 0.130) and only trivially reduces the matched-shared-trial gap from $+21.1$ to $+18.9$ pp. The actual failure mechanism is the learning rate itself: at $\eta \approx 2 \times 10^{-2}$, the very first optimization step drives an entire ReLU readout layer negative, permanently halting all subsequent gradient flow. The complex model survives this shared trial simply because its classification readout operates on $|z|$, which is strictly non-negative and inherently possesses no dead region.

Table 3: Decomposition of the complex advantage into independent constraint and datatype components. Measured on geometry-controlled seven-class RadioML, utilizing CReLU across 16 trials $\times$ 5 seeds. Seed standard deviations are provided in parentheses. The capacity-controlled benchmark for evaluating the constraint isolates against the parameter-matched real baseline.

Family	**Matched-shared-trial**	**Independent**
complex	0.439 (0.012)	0.439
real, $U(1)$-constrained	**0.448** (0.007)	**0.448**
real polar	0.433 (0.017)	0.443
real magnitude	0.395 (0.016)	0.394
real, parameter-matched	0.337 (0.021)	0.424
real stacked	0.332 (0.030)	0.402
Constraint (constrained − param-matched)	+11.1 pp	+2.4 pp
Datatype (complex − constrained)	−0.9 pp	−0.9 pp

The baseline-set confound infects the original findings. This exact three-class configuration perfectly demonstrates the overarching pattern summarized in Table 1: evaluated against the Cartesian-only baseline set, the matched-shared-trial gap is a massive $+21.1$ pp. However, simply deploying the polar real baseline at that exact same trial slashes the apparent gap to just $+5.4$ pp. When evaluated under proper independent per-family selection, the gap practically vanishes to $+0.6$ pp. The polar real representation proves vastly more stable than stacked real at the selected learning rate (0.049 standard deviation versus 0.134).

4.4 The Impact Stems from the Constraint, Not the Datatype

Proposition 1 explicitly asserts that a complex layer operates as a $U(1)$-constrained real layer. This mathematical assertion is directly measurable in practice. We integrate a fourth analytical arm into our evaluation sweep: a real-valued neural network where every individual convolution kernel is strictly restricted to the equivariant subspace $aI + bJ$, containing trainable real parameters a, b, but is otherwise identical in datatype, overall depth, internal channel layout, activation function, and optimizer to the standard real baselines. Because its internal width is parameterized in complex channels, it perfectly matches the complex model's parameter count by construction ($59,399$ against $59,406$). This novel setup cleanly separates two variables that are historically conflated: comparing constrained-real against unconstrained-real isolates the pure symmetry constraint while holding the datatype firmly fixed, whereas comparing complex against constrained-real explicitly isolates the datatype while holding the symmetry constraint fixed. This decomposition yields a remarkably one-sided conclusion (Table 3). The isolated $U(1)$ constraint provides a $+11.1$ pp boost under matched selection and a $+2.4$ pp boost under robust independent selection. Adding the complex datatype on top of that structural constraint yields an effect of -0.9 pp: the $U(1)$-constrained real network actually performs marginally *better* than the pure complex network, simultaneously proving to be the most seed-stable family across the entire experimental sweep (exhibiting a standard deviation of just 0.007, compared to 0.012 for the complex network and 0.030 for standard stacked real). While the two architectures are not identical end-to-end, the complex model classifies relying on $|z|$ whereas the constrained real model leverages a linear readout, this final residual difference remains within a single percentage point. We conclude that whatever performance gap survives the strict confounds outlined in Section 4 is wholly attributable to the structural symmetry constraint that complex multiplication naturally enforces, and that identical constraint is trivially available to any real-valued network engineered to state it directly.

5 Convergence and Activation Confounds

5.1 The Impact of Training Budget

The synthetic RF representation stress tests heavily cited in previous work were originally evaluated using an exceptionally short computational budget: specifically, three seeds, a mere 120-200 steps,

Table 4: Performance results for synthetic RF representation stress tests comparing the short and converged training budgets. Accuracy reflects the best-performing family's mean test accuracy, and the family achieving it is explicitly named. A full five of eight categorical winners completely change; in four of those specific reversals, the short budget improperly credits the complex model for a transient margin that full convergence eliminates entirely.

Condition	Short-budget (best)	acc	Converged (best)	acc	
PSK-only	complex	0.821	FLOP-matched real	0.917	$*$
QAM-only	magnitude	0.524	magnitude	0.609	
Mixed PSK+QAM	complex	0.507	FLOP-matched real	0.706	$*$
Low-SNR PSK	param-matched real	0.526	stacked real	0.570	$*$
High-SNR PSK	complex	0.949	FLOP-matched real	0.997	$*$
Unit-magnitude mixed	phase	0.491	complex	0.584	$*$
Fixed-rotation PSK	magnitude	0.328	magnitude	0.337	
Rotation-augmented PSK	complex	0.654	complex	0.725	

and a hidden width of 16. The exact same software suite includes a "converged" preset configuration – five random seeds, 600 complete steps, and a hidden width of 32, utilizing four times the underlying data, which requires approximately twenty minutes to execute on a single consumer-grade GPU. Executing this converged setup completely changes the final reported winner in five out of the eight test conditions (Table 4).

The magnitude of the margin shift matters even more than the winner flipping. On PSK-only data, historically the flagship result supposedly proving "phase-dominant data profoundly favors complex architectures", the complex model easily defeats the standard stacked-real baseline by 9.3 pp when artificially restricted to the short budget (0.821 versus 0.728). However, at convergence, this massive gap collapses to just 0.4 pp (0.9152 versus 0.9115), and ultimately, both architectures are cleanly beaten by a standard FLOP-matched real baseline reaching 0.9165. It becomes evident that nearly the entirety of the apparent architectural advantage was merely a difference in how rapidly the two competing families approach their respective asymptotes, not a fundamental difference in their representational capacity. This structural flaw perfectly mirrors the issue identified in Section 4, merely substituting a different experimental confound: an architectural performance gap that immediately vanishes the moment proper evaluation controls are applied.

5.2 Activation Choice: A Massive Effect with Task-Dependent Directions

The CReLU and ZReLU activation functions are intrinsically magnitude-blind: they operate strictly on independent real and imaginary components, or on quadrants, but they never explicitly compute or utilize $|z|$. In stark contrast, ModReLU, Siglog, and ComplexCardioid dynamically gate on magnitude. Returning to Proposition 2, every complex activation must mathematically sacrifice something; empirically, the interaction with magnitude is the axis that drives massive real-world performance differences. When evaluating EEG tasks, the mere choice of activation function shifts overall accuracy by an astonishing 58 pp on amplitude-event detection and 61 pp on complex phase-amplitude coupling. These margins are substantially larger than any standalone architectural structural effect reported anywhere else in this entire paper. Consequently, any CVNN evaluation that rigidly fixes a single activation function across varying tasks will inevitably generate conclusions like "complex models structurally fail on amplitude-based tasks," when the scientifically accurate statement is simply "the specific chosen activation function is magnitude-blind."

Magnitude-gating should be a diagnostic, not a blind prescription. The intuitive reading of the EEG amplitude-event results that $|z|$-gated activations should always be preferred whenever amplitude matters, fails to survive testing in the other two domains. We explicitly highlight this non-replication because it significantly sharpens our overall claim rather than weakening it. While ModReLU and Siglog easily beat CReLU by 13.4 and 13.0 pp on EEG amplitude events, they proceed to lose to it in every other tested scenario: they fall behind by 31.6 and 16.6 pp on phase-amplitude coupling, by 3.3 and 2.3 pp on PSK, by 2.4 and 3.0 pp on QAM, and by 5.5 and 8.2 pp

Table 5: Complex-model classification accuracy mapped by activation function, measured at the converged budget, across three separate domains. For the EEG tasks (amplitude-event detection and phase-amplitude coupling), chance is 0.25, while the polar real reference achieves 0.994 and 1.000. For RF, synthetic PSK-only and QAM-only representation tests are used. For Quantum, it is a potential inverse from a wavefunction. The highest performer per column is bolded.

Activation	EEG amp.	EEG PAC	RF PSK	RF QAM	Quantum
`ZReLU`	0.365	0.287	0.911	0.511	0.901
`CReLU`	0.814	0.578	0.925	**0.527**	**0.928**
`ComplexCardioid`	0.825	**0.871**	**0.929**	0.527	0.925
`Siglog`	0.944	0.412	0.901	0.497	0.847
`ModReLU`	**0.948**	0.262	0.892	0.503	0.874

on the quantum task. Outside the isolated EEG domain, the total performance swing across different activations is tightly bound between 3-8 pp. By design, gating on $|z|$ forces access to amplitude specifically by discarding the delicate phase-equivariant structure of the underlying layer; therefore, it only provides a benefit when the ground-truth label relies on a slow amplitude envelope and the phase acts purely as nuisance noise. ComplexCardioid, which cleverly gates on phase while passing magnitude through unaltered, remains the only evaluated activation to never fall more than 5 pp behind CReLU on any task, while concurrently outperforming it by a massive 29 pp on coupling. As such, it represents the most defensible baseline default. The rigorous practical recommendation is to explicitly report an activation's magnitude dependence as a formal experimental control, rather than universally adopting magnitude-gated activations.

The performance deficit is strictly not a tuning artifact. Before officially attributing the complex model's extreme weakness on amplitude-structured EEG data entirely to the activation function, we rigorously ruled out the learning rate as a confound. Sweeping the learning rate over a massive $100\times$ range at the short budget traps the complex model entirely between 0.317 and 0.388 on amplitude-event detection across all settings (0.0003 through 0.03). Over that exact same range, the polar real baseline dramatically swings from 0.705 up to 0.958 and back down to 0.830. For phase-amplitude coupling, the complex model remains rigidly bound between 0.237 and 0.340, while the polar baseline easily achieves a peak of 0.958. Absolutely no learning rate adjustment rescues the complex architecture. What successfully rescues it is swapping to a deliberately magnitude-aware activation function. This definitively proves why Table 5 serves as the correct diagnostic tool and a generic tuning sweep does not.

6 Related Work

Complex-Valued Architectures. The mathematical foundations of complex-valued learning Hirose (2006) and multi-valued neurons Aizenberg (2011) precede the modern deep CVNN formulation (including complex convolutions and activations) pioneered by Trabelsi et al. (2017). While these architectures report consistent gains across various domains, we propose no novel architecture; instead, we systematically isolate the exact mechanisms driving these existing performance gains.

Symmetry and Group-Equivariant Learning. Group-equivariant deep learning natively encodes task symmetries. Proposition 1 represents a one-dimensional $U(1)$ specialization of this broader principle, serving as background context. Our addition is strictly empirical: Section 4.4 rigorously separates and measures the independent effects of this mathematical constraint versus the raw complex datatype.

Theoretical Separations Between Complex and Real Neurons. Wu et al. (2023) and Wu et al. (2026) analytically prove that finite-width real networks cannot learn single complex neurons via gradient descent. This complements our thesis: they focus on isolated neuron learnability, whereas Proposition 1 maps layer-wide expressivity, showing complex layers are a $U(1)$-equivariant subset of real two-channel layers. Section 4.4 aligns with both assertions empirically, showing that holding

the symmetry constraint fixed while altering only the datatype costs just 0.9 pp. Thus, the theoretical reachability gap yields minimal empirical advantage; the mathematical constraint itself is the true engine of performance.

Tuning Protocol and Benchmarking Artifacts. Apparent architectural gains frequently shrink under rigorous testing. Dodge et al. (2019) advocate reporting validation performance by search budget (subsuming our selection rules), paralleling the optimizer benchmarking arguments of Sivaprasad et al. (2020). We contribute by demonstrating that matched-shared-trial selection exhibits a profound *directional* bias favoring optimization-robust families, driven by a clearly identified first-step mechanism.

7 Discussion and Limitations

This study successfully tests authentic real-world QAM and ASK classes, moving definitively beyond purely synthetic datasets. The pilot experiments are rigorously executed to full convergence. The identified crossover phenomenon is strictly controlled for both internal class composition and external sequence length. Finally, Proposition 1 has transitioned from a theoretical argument into an empirical measurement, as the $U(1)$-constrained real arm explicitly quantifies the constraint separately from the datatype (Section 4.4).

There are several open boundaries: **(1) Scale & Architecture:** We leverage compact 1-D convolutions on data subsets; evaluating full archives, 2-D complex convolutions, attention frameworks, or massive MRI-scale architectures remains future work. **(2) Invariance:** While Proposition 1 strictly bounds linear layers, incorporating nonlinearities and readouts means end-to-end networks lack automatic phase-invariance, as fixed-rotation stress tests confirm. **(3) Generality:** The SNR crossover is solidly established in RF, but mathematically analogous boundaries within MRI or EEG remain untested. **(4) Activation Choice:** The 58 pp accuracy swing is an EEG-specific result; in RF and quantum domains, $|z|$-gating is frequently harmful. Thus, reporting magnitude dependence is mandatory, but we cannot prescribe a universal gate. **(5) Readout Confound:** Because the complex model classifies directly from $|z|$ while the constrained real model uses a linear layer, the -0.9 pp residual strictly bounds, rather than perfectly isolates, the datatype effect. A modulus readout on the real network would close this gap.

Actionable Recommendations. Valid comparisons between CVNNs and real networks must report coordinate-view baselines (specifically polar representations for stability control), evaluate both selection rules, provide convergence evidence, and clearly document activation magnitude-dependence. To prove that complex arithmetic inherently provides an advantage, introducing a single-layer $U(1)$-constrained real arm directly and immediately settles the claim.

AI use statement

We used AI as a research assistant at a few stages of this work. It was used to *provide feedback and reproduce experiments on multiple seeds* stated in Section 5. It was used to *implement methods*: the `--grad-clip-norm` option, the global activation override in the stress-test harness, and the `real_equivariant` model family were drafted by the assistant. In the recommended categories, it was used for literature search based on the keywords from the original idea and hypothesis, and for drafting and editing portions of the manuscript. It was *not* used to generate any dataset used in the experiments, all of which are either the public RadioML archive or produced by deterministic generators that predate this work; it was *not* used to produce the theoretical content, as Propositions 1 and 2 are restatements of known results; and it did not author the underlying research codebase. All model-written code was reviewed by the authors and is covered by the project's linting, static type checking, and test suite (148 tests); every numerical claim in the paper is traceable to a committed result manifest rather than to model output; the analytical claim about the first Adam step was verified numerically as well as derived; and the references that the assistant surfaced were checked against their original sources. The authors take full responsibility for the contents of this submission.

ETHICS STATEMENT

This work raises no human-subjects concerns. The neuroscience and physics studies are representation probes on *synthetic* data: the EEG pilot generates complex analytic signals from a parametric model of phase locking, amplitude bursts, and phase-amplitude coupling rather than using any recording from a human participant, and the quantum pilot generates one-dimensional wavefunctions numerically. No clinical, biometric, personal, or otherwise identifying data was collected, processed, or released, and no institutional review was required. The one real dataset, RadioML 2018.01A, is a publicly distributed benchmark of machine-generated and over-the-air radio signals that contains no personal information; we use it under its distributor's license and do not redistribute it, providing acquisition instructions instead. We note one dual-use consideration for completeness: automatic modulation classification is relevant to spectrum monitoring and signal intelligence as well as to civilian communications. We judge the marginal risk here to be negligible, because the paper introduces no new architecture or capability, and its principal empirical finding is that previously reported advantages in this setting are substantially smaller than believed once evaluation confounds are controlled. The intended effect of this work is on evaluation practice rather than on capability: it argues that architectural comparisons in this literature should report coordinate-view baselines, convergence evidence, activation dependence, and both hyperparameter-selection rules before a gap is interpreted as architectural. Finally, the experiments were deliberately scoped to run on a single consumer GPU, so the compute and associated energy cost of this study are modest.

REPRODUCIBILITY STATEMENT

Every quantitative claim in this paper is backed by a committed result manifest. Anonymized source code for all models, baselines, experiment drivers and plotting scripts is provided in the supplementary material, and the experiment outputs are included as a `results/` tree containing one directory per reported configuration. Each directory holds a `manifest.json` conforming to the JSON schema in `docs/result_manifest.schema.json`, recording the git commit and a dirty-tree flag, the Python, PyTorch and CUDA versions, device and dtype, the complete experiment configuration, the list of seeds, and the resulting metrics, alongside the per-seed raw runs and the aggregated summary from which the reported numbers are computed. The appendix tables are generated directly from those manifests by `scripts/build_appendix_tables.py` and the crossover figure by `scripts/plot_crossover_figure.py`, so the numbers in the paper cannot drift from the runs that produced them. The complete configuration of every run, which fully determines its command line, is recorded in that run's manifest, and the runs are inventoried in Table **??**; all runs use explicit fixed seeds, and the correctness of the layers and baselines is covered by a test suite of 148 tests run in continuous integration, including an assertion that the $U(1)$-constrained real convolution matches the complex convolution in both the forward pass and gradients to 10^{-6}. Two caveats are worth stating plainly. RadioML 2018.01A is access-gated and cannot be redistributed, so Appendix D documents both how to obtain the archive and the signal-statistics procedure we used to verify that the class-index ordering is correct, since a silently permuted label order would invalidate every result derived from it. And exact numerical reproduction requires a CUDA build of PyTorch compatible with the host driver; the environment block of each manifest records the build actually used.

REFERENCES

Igor Aizenberg. *Complex-valued neural networks with multi-valued neurons*, volume 353. Springer, 2011.

Martin Arjovsky, Amar Shah, and Yoshua Bengio. Unitary evolution recurrent neural networks. In *International conference on machine learning*, pp. 1120–1128. PMLR, 2016.

Taco S Cohen and Max Welling. Steerable cnns. *arXiv preprint arXiv:1612.08498*, 2016.

Elizabeth Cole, Joseph Cheng, John Pauly, and Shreyas Vasanawala. Analysis of deep complex-valued convolutional neural networks for mri reconstruction and phase-focused applications. *Magnetic resonance in medicine*, 86(2):1093–1109, 2021.

Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, and Noah A Smith. Show your work: Improved reporting of experimental results. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2185–2194, 2019.

Akira Hirose. *Complex-valued neural networks*. Springer, 2006.

Timothy James O'Shea, Tamoghna Roy, and T. Charles Clancy. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):168–179, 2018. doi: 10.1109/JSTSP.2018.2797022.

Prabhu Teja Sivaprasad, Florian Mai, Thijs Vogels, Martin Jaggi, and François Fleuret. Optimizer benchmarking needs to account for hyperparameter tuning. In *International conference on machine learning*, pp. 9036–9045. PMLR, 2020.

Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep complex networks. *arXiv preprint arXiv:1705.09792*, 2017.

Patrick Virtue, X Yu Stella, and Michael Lustig. Better than real: Complex-valued neural nets for mri fingerprinting. In *2017 IEEE international conference on image processing (ICIP)*, pp. 3953–3957. IEEE, 2017.

Scott Wisdom, Thomas Powers, John Hershey, Jonathan Le Roux, and Les Atlas. Full-capacity unitary recurrent neural networks. *Advances in neural information processing systems*, 29, 2016.

Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5028–5037, 2017.

Jin-Hui Wu, Shao-Qun Zhang, Yuan Jiang, and Zhi-Hua Zhou. Complex-valued neurons can learn more but slower than real-valued neurons via gradient descent. *Advances in Neural Information Processing Systems*, 36:23714–23747, 2023.

Jin-Hui Wu, Shao-Qun Zhang, Yuan Jiang, and Zhi-Hua Zhou. Reparameterized complex-valued neurons can efficiently learn more than real-valued neurons via gradient descent. *Journal of Machine Learning Research*, 27(38):1–51, 2026.

A PROOFS

Proof of Proposition 1. We directly demonstrate the mathematical equivalence by proving (i) $\Leftrightarrow$ (ii) and (ii) $\Leftrightarrow$ (iii).

(ii) $\Rightarrow$ (i). Assuming $W = aI + bJ$, we observe that W functions directly as a polynomial in J. Since the rotation operator $R_\phi = \exp(\phi J)$ constitutes a convergent power series entirely in J, the two terms inherently commute.

(i) $\Rightarrow$ (ii). Assuming $WR_\phi = R_\phi W$ for all values of ϕ. Differentiating this relationship at $\phi = 0$ yields $WJ = JW$. Expanding both sides using the notation $W = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$, we derive:

$$WJ = \begin{pmatrix} q & -p \\ s & -r \end{pmatrix}, \qquad JW = \begin{pmatrix} -r & -s \\ p & q \end{pmatrix}.$$

This strictly implies $q = -r$ and $p = s$. Consequently, W must take the precise form $pI + qJ$.

(ii) $\Leftrightarrow$ (iii). Utilizing the standard mathematical embedding of $\mathbb{C}$ into $\mathbb{R}^{2\times 2}$ that explicitly maps $x + iy$ to the vector $(x, y)^\top$, the standard complex multiplication by $a + ib$ effectively maps $x + iy$ to $(ax - by) + i(ay + bx)$. This transformation operates identically to the matrix $aI + bJ$. This defined correspondence represents a complete bijection between the set $\{aI + bJ : a, b \in \mathbb{R}\}$ and the complex plane $\mathbb{C}$, confirming it acts as a formal ring isomorphism. $\square$

The established two-parameter family $aI + bJ$ physically resides inside the broader four-parameter space defining unconstrained real 2×2 channel mixings; therefore, the complex layer acts as a strict codimension-two subspace of the standard stacked-real layer, never an extension of it.

Proof of Proposition 2. Liouville's underlying theorem formally dictates that any function on $\mathbb{C}$ that is both bounded and entire must be constant. Therefore, any proposed activation σ attempting to be simultaneously bounded, mathematically entire, and nonconstant directly violates this established theorem. □

B WIRTINGER CALCULUS

By explicitly defining $f : \mathbb{C} \to \mathbb{C}$ in the expanded form $f(z) = u(x, y)+iv(x, y)$ where $z = x+iy$, we establish the standard Wirtinger derivatives:

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\,\frac{\partial f}{\partial y}\right), \qquad \frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\,\frac{\partial f}{\partial y}\right).$$

The standard Cauchy-Riemann equations fundamentally equate to the condition $\partial f/\partial \bar{z} = 0$; therefore, strict holomorphy defines the exact specific scenario where the conjugate derivative completely vanishes. Standard backpropagation explicitly does not mandate this condition. When optimizing a standard real-valued loss $L : \mathbb{C}^n \to \mathbb{R}$ targeting a complex parameter w, the algorithm performs steepest descent directly in the underlying real plane $(\Re w, \Im w)$. This descent trajectory is perfectly captured by the conjugate Wirtinger derivative. The mathematical update step $w \leftarrow w - \eta\, \overline{\partial L/\partial \bar{w}}$ represents the precise convention implemented by modern autodiff frameworks to handle complex parameters natively.

Two specific mathematical consequences are highly relevant here. First, activation functions gating explicitly on $|z|$—such as ModReLU, Siglog, and ComplexCardioid—remain structurally differentiable almost everywhere when treated as mapping functions $\mathbb{R}^2 \to \mathbb{R}^2$. Consequently, they remain entirely valid, trainable components of a neural network even if they are technically nowhere holomorphic. Second, the resulting Cauchy-Riemann residual $|\partial\sigma/\partial\bar{z}|$ functions not as a hard constraint, but rather as a highly practical *diagnostic* metric: it directly quantifies precisely how far a chosen activation deviates from pure holomorphy, a metric Appendix C explicitly tracks.

C ACTIVATION GALLERY AND CAUCHY-RIEMANN RESIDUALS

Table 6 comprehensively characterizes the behavior of six prominent complex activations across a standardized grid spanning $z \in [-3, 3]^2$, programmatically generated by our repository's `characterize_activations` script. The results make the theoretical Liouville trade-off (outlined in Proposition 2) heavily concrete: no evaluated activation manages to remain simultaneously bounded and perfectly holomorphic. Two specific entries demand explicit commentary. ZReLU's median residual effectively registers as exactly zero because the function behaves as a pure identity mapping on one specific quadrant while resolving to zero elsewhere; consequently, it proves locally holomorphic almost everywhere. Its entire residual mathematical mass aggressively concentrates directly on the rigid quadrant boundaries, a phenomenon clearly documented by the massive max residual recorded. Meanwhile, Siglog remains the absolute only bounded activation evaluated in the entire set (with $\max|\sigma| = 0.809$). It directly pays for this mathematical boundedness by incurring the smallest Jacobian norm in the entire test group by a full order of magnitude—representing the stability-versus-gradient-flow consequence of the trilemma, physically manifested as measurable data.

D RADIOML CLASS-ORDER VERIFICATION

Table 7 formally documents the exact index-to-modulation map heavily utilized in our study, paired directly with the empirical statistics validating it. All metrics were gathered based on 256 examples per class rigorously evaluated at 30,dB, with every individual example mathematically normalized to target unit mean power. The "Near-0" metric explicitly captures the raw fraction of samples

Table 6: Detailed activation behavior characterized across the bounding grid $z \in [-3, 3]^2$. The "CR" metric denotes the Cauchy–Riemann residual $|\partial\sigma/\partial\bar{z}|$, which registers at exactly zero in regions where the underlying activation is locally holomorphic. While ComplexTanh acts as the sole purely holomorphic candidate, it remains inherently unbounded: the test grid deliberately brackets its critical poles at $\pm i\pi/2$, driving its residual median down near absolute zero while causing its max and $\max|\sigma|$ to explosively scale. In contrast, every strictly bounded activation inherently pays a continuous nonzero residual penalty.

Activation	$\max\|\sigma\|$	**CR (median)**	**CR (p95)**	**CR (max)**	**Jacobian norm**
CReLU	4.243	0.5	1	1	2.65
ZReLU	4.243	0	2.2e−15	29.5	1.16
ModReLU	3.993	0.105	0.316	0.877	4.08
ComplexCardioid	3.621	0.353	0.497	0.500	1.95
Siglog	0.809	0.207	0.248	0.250	0.142
ComplexTanh	48.08	0.000578	0.0633	905.4	108

Table 7: The formal RadioML 2018.01A one-hot index physically mapped directly to specific modulation, supported entirely by validating statistical checks run at 30,dB. The amplitude metrics perfectly separate all amplitude-bearing signal classes, cleanly identify OOK specifically by analyzing its near-zero fraction, and successfully identify FM and GMSK exactly as constant-envelope. The rigorous M-th power test definitively confirms BPSK and QPSK. However, at this designated sequence length, the test mathematically fails to resolve 8PSK and higher modulation orders.

#	**Modulation**	**CV**	**Near-0**	#	**Modulation**	**CV**	**Near-0**
0	OOK	0.767	11.1%	12	16QAM	0.407	0.5%
1	4ASK	0.634	7.8%	13	32QAM	0.408	0.5%
2	8ASK	0.579	5.4%	14	64QAM	0.419	0.6%
3	BPSK	0.462	2.2%	15	128QAM	0.413	0.6%
4	QPSK	0.311	0.3%	16	256QAM	0.423	0.6%
5	8PSK	0.296	0.3%	17	AM-SSB-WC	0.039	0.0%
6	16PSK	0.299	0.3%	18	AM-SSB-SC	0.072	0.0%
7	32PSK	0.299	0.3%	19	AM-DSB-WC	0.492	6.3%
8	16APSK	0.371	0.4%	20	AM-DSB-SC	0.498	6.3%
9	32APSK	0.452	0.7%	21	FM	0.005	0.0%
10	64APSK	0.409	0.5%	22	GMSK	0.085	0.0%
11	128APSK	0.422	0.5%	23	OQPSK	0.391	1.2%

whose isolated magnitude drops completely below one tenth of that specific example's rolling mean magnitude. The critical M-th power column accurately records the core constellation order cleanly recovered directly from the strict peak-to-median ratio of $\mathrm{FFT}((z/|z|)^M)$. The column remains explicitly blank wherever no decisive order could be confidently established.

E Quantum wavefunction pilot

The quantum pilot probes whether the representation logic transfers outside RF. Momentum is carried by the phase gradient and is invisible to $|\psi|$, so every phase-aware view solves it and the magnitude-only view stays at chance; the potential-inverse and global-phase tasks mix amplitude and phase structure, and there the explicit polar view wins. These are representation probes, not domain benchmarks.

Table 8: Quantum wavefunction pilot at the converged budget. Chance is 0.25 for momentum and 0.20 for the potential tasks. Columns are the complex model and the stacked-real, phase-only, polar and magnitude-only real baselines.

Condition	complex	stacked	phase	polar	magnitude
momentum_phase	1.0000	1.0000	1.0000	1.0000	0.2500
potential_inverse	0.8863	0.8894	0.8792	0.9227	0.6749
global_phase_shift	0.8482	0.8353	0.8322	0.8800	0.6733
global_phase_augmented	0.8922	0.8827	0.8365	0.8980	0.6773

Table 9: EEG analytic-signal pilot at the converged budget, CReLU. Chance is 0.25 throughout. Columns are the complex model and the stacked-real, phase-only, polar and magnitude-only real baselines.

Condition	complex	stacked	phase	polar	magnitude
phase_locking	0.9985	0.9995	0.9995	1.0000	0.2500
amplitude_event	0.8137	0.8225	0.2456	0.9936	1.0000
phase_amplitude_coupling	0.5779	0.8510	0.2475	1.0000	0.2500
reference_phase_shift	0.9975	0.9995	1.0000	1.0000	0.2500
reference_phase_augmented	0.9995	1.0000	0.9990	0.9995	0.2500

F EEG ANALYTIC-SIGNAL PILOT, FULL TABLES

The EEG pilot separates phase locking, amplitude events and phase–amplitude coupling on synthetic analytic signals. Phase-dominant conditions are solved by every phase-aware view; the amplitude and coupling conditions are the ones that discriminate, and they are where the complex model's activation choice dominates (Table 5).

Table 10: EEG learning-rate control at the short budget. The complex model's deficit on both amplitude-structured tasks is stable across a $100\times$ range, so it is not a tuning artifact.

	Amplitude event		Phase–amp. coupling	
Learning rate	**complex**	**polar real**	**complex**	**polar real**
0.0003	0.3173	0.7051	0.2372	0.4327
0.001	0.3205	0.9519	0.2788	0.7276
0.003	0.3173	0.9583	0.2821	0.9038
0.01	0.3878	0.9455	0.3397	0.9583
0.03	0.3814	0.8301	0.3109	0.4423

G PER-SNR TABLES, ALL CONFIGURATIONS

Figure 2 plots the first and fourth of these. All four are given here so the crossover band can be read off directly: in every configuration the complex row leads the polar row below 2 dB and trails it above 6 dB.

Table 11: Seven classes, length 128, CReLU. Matched-shared-trial selection, mean over seeds.

Family	SNR (dB)								
	−14	−10	−6	−2	2	6	10	14	18
complex	0.158	0.160	0.265	0.385	0.534	0.594	0.618	0.614	0.604
stacked real	0.136	0.138	0.168	0.262	0.386	0.482	0.479	0.487	0.490
param-matched real	0.146	0.137	0.177	0.263	0.376	0.431	0.449	0.452	0.447
FLOP-matched real	0.149	0.138	0.161	0.276	0.426	0.504	0.526	0.521	0.517
polar real	0.138	0.147	0.161	0.288	0.482	0.632	0.674	0.670	0.676
phase-only real	0.137	0.141	0.149	0.172	0.260	0.278	0.271	0.273	0.276
magnitude-only real	0.141	0.138	0.174	0.255	0.419	0.566	0.631	0.628	0.629

Table 12: Seven classes, length 128, ZReLU. Matched-shared-trial selection, mean over seeds.

Family	SNR (dB)								
	−14	−10	−6	−2	2	6	10	14	18
complex	0.158	0.151	0.243	0.398	0.504	0.546	0.562	0.567	0.560
stacked real	0.138	0.134	0.162	0.241	0.379	0.474	0.475	0.477	0.479
param-matched real	0.144	0.132	0.152	0.255	0.405	0.454	0.476	0.480	0.470
FLOP-matched real	0.146	0.142	0.160	0.269	0.416	0.492	0.509	0.498	0.486
polar real	0.143	0.141	0.171	0.273	0.476	0.610	0.663	0.663	0.659
phase-only real	0.132	0.144	0.149	0.179	0.265	0.286	0.270	0.275	0.284
magnitude-only real	0.142	0.149	0.162	0.267	0.423	0.566	0.612	0.638	0.634

Table 13: Ten classes, length 128, CReLU. Matched-shared-trial selection, mean over seeds.

Family	SNR (dB)								
	−14	−10	−6	−2	2	6	10	14	18
complex	0.098	0.125	0.174	0.267	0.370	0.425	0.425	0.428	0.437
stacked real	0.092	0.096	0.110	0.165	0.256	0.287	0.287	0.294	0.293
param-matched real	0.100	0.103	0.108	0.132	0.210	0.228	0.233	0.231	0.227
FLOP-matched real	0.105	0.101	0.110	0.153	0.256	0.275	0.273	0.274	0.279
polar real	0.107	0.097	0.110	0.190	0.351	0.434	0.468	0.448	0.467
phase-only real	0.100	0.100	0.100	0.112	0.117	0.135	0.139	0.138	0.137
magnitude-only real	0.103	0.096	0.108	0.165	0.293	0.383	0.415	0.418	0.434

Table 14: Seven classes, length 256, CReLU. Matched-shared-trial selection, mean over seeds.

Family	SNR (dB)								
	−14	−10	−6	−2	2	6	10	14	18
complex	0.162	0.201	0.308	0.416	0.575	0.624	0.642	0.637	0.629
stacked real	0.149	0.149	0.176	0.300	0.468	0.520	0.518	0.523	0.532
param-matched real	0.144	0.142	0.159	0.270	0.425	0.490	0.501	0.484	0.501
FLOP-matched real	0.141	0.139	0.167	0.253	0.451	0.526	0.521	0.530	0.530
polar real	0.144	0.146	0.185	0.334	0.569	0.703	0.718	0.730	0.727
phase-only real	0.150	0.151	0.158	0.189	0.260	0.304	0.315	0.319	0.317
magnitude-only real	0.152	0.133	0.163	0.274	0.482	0.638	0.676	0.685	0.697