FROM SPECTRAL RESPONSES TO TASK EFFECTS: AUDITING FREQUENCY DEPENDENCE IN SPIKING NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Connecting a spectral response in a spiking neural network to task behavior requires measurements beyond band energy. We present a frequency-lineage protocol that separates operator responses, fixed-decoder responses, and predictive interventions. A MaxFormer trace locates the spread of an input-band perturbation at the first LIF, and a three-architecture, three-dataset panel records cross-frequency decoder responses. Matched operator experiments place this observation in context: ordinary ReLU and tanh also generate cross-frequency components, while integration and reset change task performance. In three frozen MaxFormer networks, equal-budget binary patches selected using same- or cross-band targets reduce classification loss, but the tested joint non-source target does not. These patches preserve binary states without strict band isolation; continuous Fourier patches give a different joint response at a larger dose. Separately, learned natural residuals show no consistent predictive advantage across the development panel, and beating a control need not mean beating the baseline. These findings motivate distinct tests for where a frequency perturbation acts, how a decoder responds, and whether an intervention benefits the task. They constrain explanations based on spectral attenuation without identifying information preservation or loss.

1 INTRODUCTION

Frequency-based SNN designs use filtering arguments to motivate which components a network should retain or enhance (Fang et al., 2025; Ye et al., 2026; Feng et al., 2025). Such arguments connect an observable quantity, spectral energy, to a practical objective, task performance. The connection requires care. Subthreshold membrane integration can attenuate high temporal frequencies, while thresholding and other nonlinear operations can distribute the effect of an input band across output frequencies. Attenuation alone therefore leaves open how the network processes task-relevant input structure. This matters when choosing what to restore.

Consider perturbing one input band and observing a response in another band after a spiking layer. This observation locates an effect. A change in a fixed destination decoder's predictions links the effect to that decoder, but the original classifier may use a different readout. Restoring a clean-reference state tests that original path under input corruption. Supplying an upstream residual on natural inputs asks a separate question: whether the residual adds predictive value beyond downstream features. These measurements have different targets, so their results must be interpreted separately.

We organize them into a frequency-lineage protocol (Figure 1) and apply it to short-window event tasks. Our contribution is the joint use of operator localization, fixed destination-band decoding, and controlled predictive interventions to audit frequency-based explanations. The experiments ask where an input-band effect spreads, whether destination decoders respond, and which interventions improve task prediction. The frequency axis indexes sequence steps rather than physical hertz.

The evidence yields concrete distinctions. In a MaxFormer trace, the first LIF spreads an input-band perturbation beyond its source band; a separate three-architecture panel finds cross-band decoder responses. Binary patches chosen by cross-band targets can improve the original classifier, yet combining non-source targets at the same budget does not yield recovery. Natural residuals provide

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.

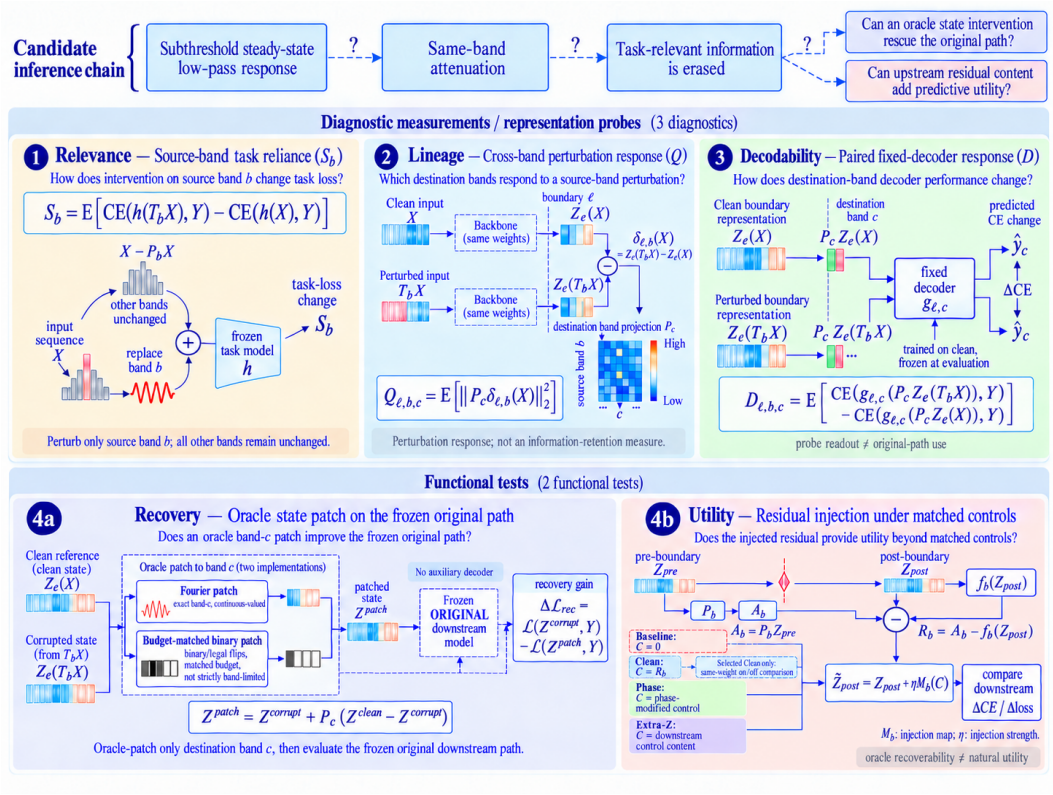


Figure 1: Three diagnostics and two complementary functional tests. Panels 1–3 measure input-band reliance, perturbation lineage, and fixed-decoder response. Panel 4a tests clean-reference recovery under source corruption; panel 4b tests learned residual utility on natural inputs. The projection formula and "only destination band" wording in 4a describe the Fourier implementation. Binary patches use band targets to select legal flips without strict band isolation. The candidate chain and diagrams are conceptual; the functional tests do not presume that natural information erasure has been established.

a further contrast: a Clean arm can have lower mean loss than content and capacity controls while losing to Baseline. These observations motivate checking predictive effects explicitly instead of assigning task value from a spectral response. Matched operators establish that mixing is not specific to spikes, and the natural experiments remain development evidence rather than a test of information conservation.

2 Related Work

Frequency-based architectural explanations. Spectral analyses of vision transformers connect attention and convolution to spatial filtering (Park & Kim, 2022; Wang et al., 2022). SNN designs include high-frequency enhancement in MaxFormer (Fang et al., 2025), video pass-band optimization (Ye et al., 2026), compartment-specific temporal filtering in TS-LIF (Feng et al., 2025), frequency selection for forecasting in SpikF (Wu et al., 2025), and redundancy suppression through spatial–temporal attention in FSTA-SNN (Yu et al., 2025). These approaches have distinct objectives. In particular, TS-LIF explicitly omits spike generation and reset from its transfer-function derivation, while FSTA's spatial module applies DCT after temporal averaging. Our experiments examine temporal processing rather than retesting spatial filtering claims. Neither spatial smoothing nor architectural improvement alone establishes temporal task-information erasure.

Neuronal dynamics and routing precedents. The distinction between membrane filtering and firing response has a long history: synaptic noise and filtering can alter high-frequency firing-rate

responses (Brunel et al., 2001), while resonate-and-fire neurons have frequency-selective dynamics (Izhikevich, 2001). Balanced RF neurons and S5-RF develop trainable resonant sequence models (Higuchi et al., 2024; Huber et al., 2025). SWAT already uses facilitating synapses to route inputs by frequency to different hidden neurons (Wade et al., 2010). Our use of routing denotes a more specific measurement: input-band intervention effects across output spectral bands and operator boundaries, coupled to task decoding. We study this behavior in MaxFormer, SEW-ResNet (Fang et al., 2021), and Spikformer (Zhou et al., 2023); we do not claim to introduce frequency-selective neuronal computation.

Spectral measurements and readable task information. Function-space spectral bias (Rahaman et al., 2019), input-frequency robustness (Yin et al., 2019), and cross-frequency statistical dependence (Malladi et al., 2018) address different quantities from our intervention-based temporal lineage. Phase can preserve signal structure absent from a magnitude-only description (Oppenheim & Lim, 1981); surrogate-data analysis also makes the preserved constraints and null hypothesis explicit (Theiler et al., 1992). For task readout, predictive $\mathcal{V}$-information and conditional probing formalize prediction relative to an observer and a baseline (Xu et al., 2020; Hewitt et al., 2021). Control tasks and MDL probing address probe learning and extraction complexity (Hewitt & Liang, 2019; Voita & Titov, 2020), whereas an information-estimation objective can favor stronger probes (Pimentel et al., 2020). We therefore use matched capacity for comparisons without treating probe failure as absence of information. Our routing scores are neither MI-in-frequency estimates nor information-conservation measurements.

Causal localization and controlled erasure. Causal abstraction and circuit analysis test explanations through interventions (Geiger et al., 2021; Wang et al., 2023), while MIB standardizes circuit and causal-variable localization evaluations (Mueller et al., 2025). Activation-patching results can depend on corruption methods and evaluation metrics (Zhang & Nanda, 2024). LEACE provides erasure guarantees for linear concept prediction (Belrose et al., 2023); such observer-specific guarantees differ from absence of task information for unrestricted prediction. Our rescue controls address whether accessible upstream content improves prediction beyond content and capacity alternatives. We claim neither a complete causal abstraction nor unrestricted erasure, and phase interventions may move representations outside the natural distribution.

3 Separating spectral responses from task effects

The frequency-lineage protocol separates measurements of a perturbation from tests of predictive benefit. Let $X \in \mathbb{R}^{T\times C\times H\times W}$ be an event-count sequence and $Z_\ell(X)$ its representation at boundary ℓ. We take the Fourier transform along sequence steps. A shared per-step linear map preserves temporal bins under fixed coefficients, and stable linear integration has a steady-state low-pass response. A finite causal window can nevertheless produce off-diagonal DFT energy through its initial-state transient, even without a nonlinearity. Thresholding and multiplicative interactions can also generate new frequencies; none of these operator capabilities alone establishes a task benefit (Appendix A). A decoder's CE change measures accessibility to that trained predictor, rather than unrestricted task information (Appendix B).

Input relevance and same-band energy. Let P_b project onto a conjugate-symmetric frequency band and $\mathcal{T}_b$ perturb its phase while preserving its Fourier magnitudes. For a fixed classifier h, define model reliance

$$S_b = \mathbb{E}[\mathrm{CE}(h(\mathcal{T}_b X), Y) - \mathrm{CE}(h(X), Y)]. \tag{1}$$

This is reliance on the perturbed structure; clean band power and its within-layer share are energy measures. DC, Nyquist, and normalization details are in Appendix D.

Where does the input-band effect appear? For each source band b and boundary ℓ, define

$$\delta_{\ell,b}(X) = Z_\ell(\mathcal{T}_b X) - Z_\ell(X), \qquad Q_{\ell,bc} = \mathbb{E}\|P_c \delta_{\ell,b}(X)\|_2^2. \tag{2}$$

We report absolute perturbation power and normalized destination shares, excluding near-zero rows from the latter. These shares describe the clean-to-perturbed difference, not attenuation of the clean representation or a fraction of task information retained. The unified panel uses a separate DC sign control (Appendix D).

A fixed-capacity decoder $g_{\ell,c}$ is fit on the clean representation of destination band c. Its paired CE response is

$$D_{\ell,bc} = \mathbb{E}[\mathrm{CE}(g_{\ell,c}(P_c Z_\ell(\mathcal{T}_b X)), Y) - \mathrm{CE}(g_{\ell,c}(P_c Z_\ell(X)), Y)]. \tag{3}$$

Off-diagonal Q locates perturbation effects; off-diagonal D measures their impact on fixed label decoders. A CE increase can also reflect collateral disruption of destination features. Neither quantity proves content transport or original-classifier use.

Two predictive tests. Original-head patches replace part of a corrupted activation with its clean-reference counterpart, keeping the downstream model fixed (Section 4.4). This tests recovery from imposed input corruption. To ask whether upstream features add value on natural inputs, we instead fit a residual side path under four matched arms. At a selected boundary let $R_b = P_b Z_{\mathrm{pre}} - f_b(Z_{\mathrm{post}})$, with f_b fit separately and then frozen. Matched Baseline, Clean, Phase, and Extra-Z systems inject zero, R_b, a phase-perturbed upstream residual, or a transformed downstream view through the same adapter. We compare task CE, accuracy, and the selected Clean model with its side path switched off at fixed weights. Phase preserves the upstream band's magnitude before reconstruction subtraction; its final residual energy need not match Clean. Full arm definitions, capacity controls, and the historical auxiliary V_G statistic appear in Appendix D and Appendix B. Validation-based selection makes these results development evidence; predictive benefit alone does not identify lost upstream information.

4 Experiments

4.1 Experimental setup

We first establish operator capabilities, then locate perturbation effects in a natural-network case and examine destination readouts across models. The two predictive experiments distinguish original-path patch recovery from learned residual utility. The unified panel crosses MaxFormer, SEW-ResNet, and Spikformer with CIFAR10-DVS (Li et al., 2017), N-Caltech101 (Orchard et al., 2015), and DVS128Gesture (Amir et al., 2017), with $T = 16$ and eight frequency groups. Groups $b_0, \ldots, b_6$ contain individual rFFT bins; b_7 contains bins 7 and 8. The unified intervention randomizes bin 7's phase but leaves Nyquist bin 8 unchanged; DC uses random signs. The shared boundary is each architecture's first continuous-to-spike stem gate, before downstream task mixing. It has 16, 64, and 32 channels, respectively. This Spikformer boundary precedes the attention block.

Each cell uses three networks and two adapter initializations per network. Class-balanced fit/select/evaluation counts are $800/200/200$, $1212/404/404$, and $847/55/55$ for CIFAR, N-Caltech, and Gesture. New-dataset evaluation images are excluded from backbone gradient training but inform checkpoint selection; historical CIFAR exposure remains. These are development results, with no new official-test evaluation. Repeated seeds do not add independent images. Appendix D records the checkpoint repairs, historical splits, and training recipes. Event-count framing (Gallego et al., 2022) makes frequency a sequence-step, rather than physical-time, quantity.

4.2 Cross-frequency generation is not specific to spikes

Before attributing a cross-frequency response to spiking dynamics, we test whether ordinary nonlinearities can generate it. In a fixed dual-tone audit, ReLU and tanh produce both difference and sum components, whereas the linear operator's corresponding power rounds to zero (Appendix Table 4). This is a capability check using one preselected fitted state per operator, not a comparison of task performance or training variability. Eight operators with the same affine architecture and training budget, but separately learned weights, then test waveform fitting. Table 1 retains the contrasts needed to separate integration, ordinary nonlinearity, thresholding, and reset; the full eight-operator matrix and protocol are in Appendix E.

Linear-IIR lowers nuisance-task error relative to Linear, while adding thresholding lowers doubling error. Adding reset raises doubling error but lowers nuisance error relative to IIR+Threshold. The same operator change thus has different consequences across targets. The spectrum audit and waveform fits separate the ability to generate a component from its use in a particular task; the next experiment asks where such responses arise in a trained event classifier.

Table 1: Selected matched-operator contrasts: extrapolation waveform MSE (lower is better), mean over three training repeats. Full uncertainties, four tasks, and eight operators appear in Appendix Table 3. The rows vary integration, pointwise nonlinearity, and subtractive reset.

Operator	Doubling	Difference	Nuisance
Linear	0.5054	0.5126	0.4051
ReLU	0.5054	0.3037	0.3519
tanh	0.5054	0.3052	0.3521
Linear-IIR	0.5054	0.5126	0.3768
IIR+Threshold	0.2833	0.3503	0.3790
LIF	0.3895	0.3531	0.3631

4.3 From the first LIF to cross-band decoder responses

The first LIF marks a transition from same-band to cross-band perturbation responses in a historical MaxFormer/CIFAR10-DVS checkpoint. Input-b_5 perturbation power remains same-band through projection and evaluation-mode BN, then spreads at the first LIF (Figure 2, top). Its same-band share falls from 100.0% to 30.6% and then 19.1% at embedding output. The fixed-decoder responses in the same trace provide a separate view of how the intervention affects label readout. The falling energy share locates the spread of a perturbation; it does not measure clean-band attenuation or show that source content has been preserved elsewhere. This case motivates the broader endpoint comparison below, whose checkpoints and protocol remain separate.

We next ask whether cross-band decoder responses also occur across architectures and datasets under the unified protocol. These panels measure the first stem spike rather than locating the transition within each architecture. Natural class labels remain readable from individual destination bands, and interventions on other source bands can change those predictions. For each of 27 networks, fixed linear destination decoders are fitted on clean features before source-band interventions. The CIFAR heatmaps in Figure 2 (bottom) show distinct AC off-diagonal patterns across architectures. Such responses also recur in the other two datasets, with task-dependent signs and magnitudes (Appendix Table 6). A near-zero signed mean can hide opposite-signed cells. These are decoder responses under intervention, potentially including distribution shift, and do not locate the internal transition in SEW or Spikformer. The six N-Caltech101 and Gesture matrices are in Appendix F.3; separate historical localization traces are in Appendix H.

4.4 Original-path recovery depends on patch domain and dose

The decoder results motivate a test on the original classifier: can clean-reference patches selected by a different destination band reverse part of the loss caused by input corruption? We freeze three MaxFormer/CIFAR10-DVS networks, including their original heads and BN statistics, and evaluate all eight source and destination bands on the same 900 development images. No head, adapter, or reconstructor is trained. At the first stem LIF output, the Fourier oracle target for a source-b corrupted pass and destination c is

$$Z^{\text{patch}} = Z^{\text{corrupt}} + P_c(Z^{\text{clean}} - Z^{\text{corrupt}}). \tag{4}$$

The matched binary intervention uses this target to rank legal corrupt-to-clean bit flips, applying the same 10% Hamming budget and clean-state distance to every single-band and joint arm. Applied flips are enriched for the requested destination, but are not strictly band-limited. A joint arm targets all non-source AC destinations in one actual forward pass. The original Fourier patch is retained as a separate continuous-value sensitivity check; its phase-randomized increment matches spectral magnitude. Zero replacement checks the corrupted path. All clean-reference patches are oracle probes, not deployable restorations.

Under the matched binary budget, patches selected by same-band and single cross-band targets yield CE recoveries of 0.0221 and 0.0115 nats; the actual joint non-source AC arm is -0.0036 nats (Figure 3; Appendix Table 11). Selecting flips by a cross-band target can therefore improve the original classifier, but the tested joint target does not yield recovery at this dose. Because binary flips are not strictly band-limited, this benefit cannot be assigned exclusively to the requested destination band. The continuous Fourier joint arm recovers 0.0243 nats, but its increment energy is about six

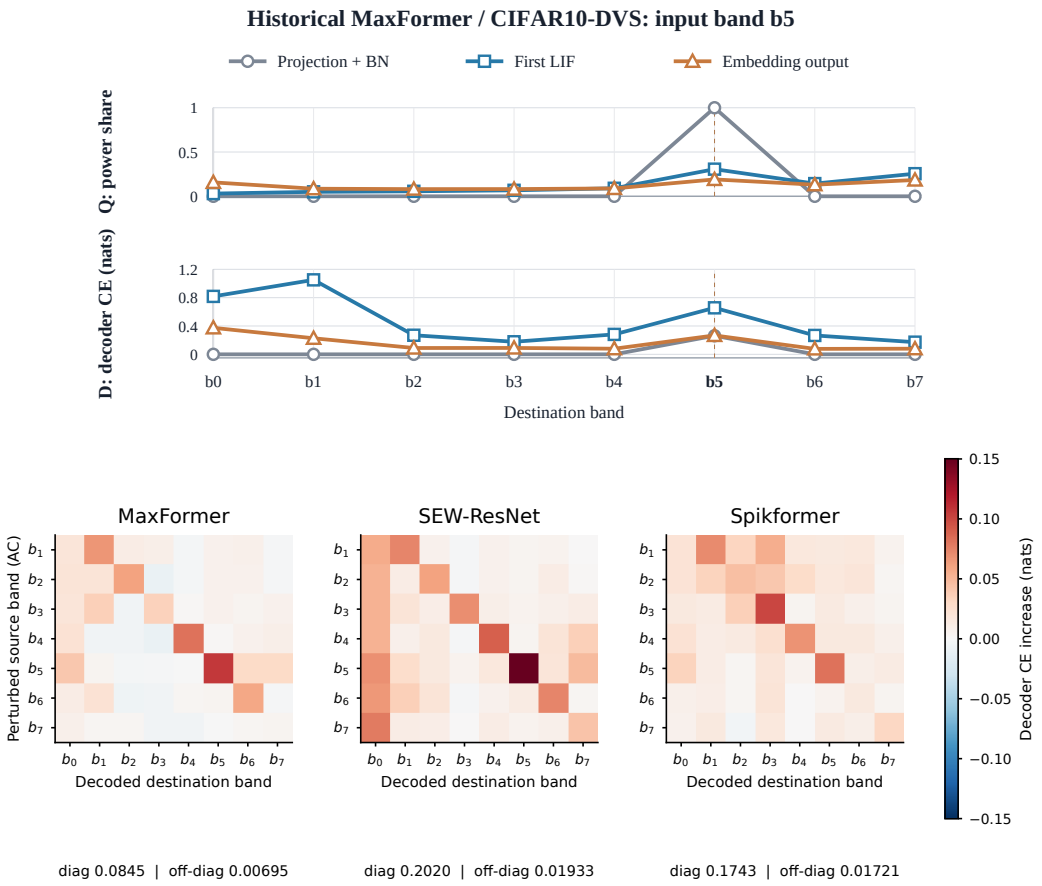


Figure 2: Localization and breadth of source-band effects. Top: historical MaxFormer input-b_5 perturbation shares and fixed-decoder CE response across projection/evaluation BN, first LIF, and embedding; one checkpoint, 9000/1000 analysis. Bottom: unified first-stem decoder CE responses for CIFAR10-DVS, with each architecture matrix averaging three networks. Heatmap rows are AC source bands, columns are destination bands, and all three matrices share a signed scale. The plotted diagonal/off-diagonal summaries include a separate DC sign control, whereas the heatmaps show AC sources. The historical trace locates a transition; the unified panels compare endpoints under a different protocol. Neither establishes information transport. The remaining dataset matrices and the three-dataset summary appear in Appendix H and Appendix F.3.

times a single cross-band patch and its state is nonbinary. This domain and dose dependence prevents a synergy or information-preservation inference. Full clean-state replacement exactly reproduces clean outputs, validating the downstream replay, while remaining an oracle operation (Appendix G).

4.5 Residual utility requires both baseline and control comparisons

Oracle recovery under source corruption leaves a separate question open: does an upstream residual improve prediction on natural inputs? We test this with learned side paths and fresh heads, without supplying a clean-reference state from a second pass. This experiment evaluates additional predictive value, rather than recovery from the imposed corruption used in Section 4.4. All 27 networks pass task-baseline health checks. Within each network, band, and adapter seed, the four arms share the frozen backbone, fresh-head initialization, sample order, and training budget. A one-epoch reconstructor fit precedes up to three epochs of head/adapter training; select CE chooses the state. The adapters have 257 parameters but differ in rank across architectures (Appendix D). A sepa-

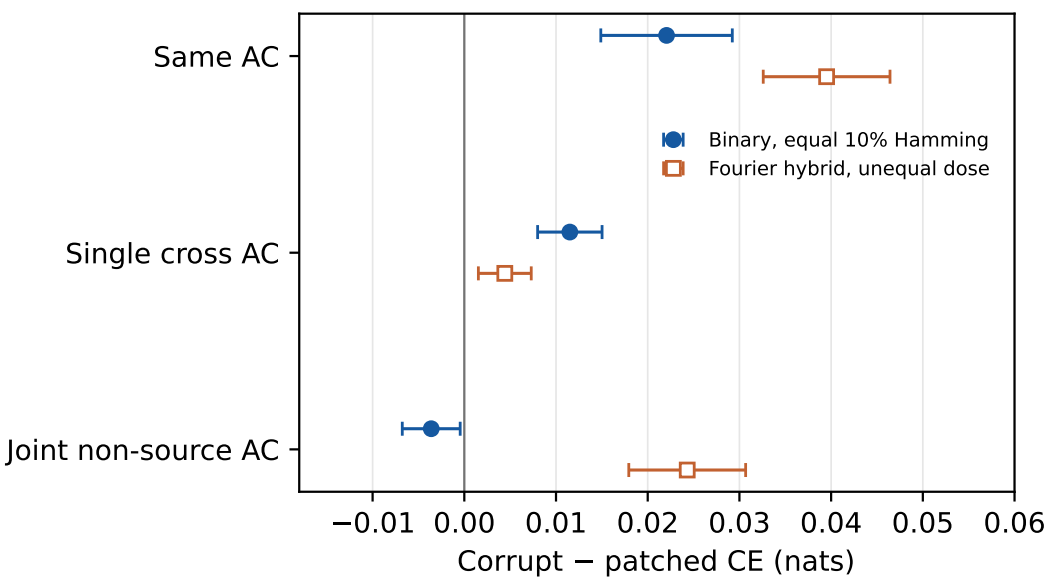


Figure 3: Original-head oracle recovery across three frozen MaxFormer networks. Points are the mean corrupted-minus-patched CE over network summaries; bars show network-level SD. Binary patches flip an equal 10% of eligible bits and preserve the post-LIF binary state. The Fourier hybrid patches inject continuous increments and have unequal doses across patch types. Each joint point comes from one forward with all non-source AC destinations, not a sum of single-band effects. DC controls are separate. Positive values favor the patch; the two intervention domains are not pooled.

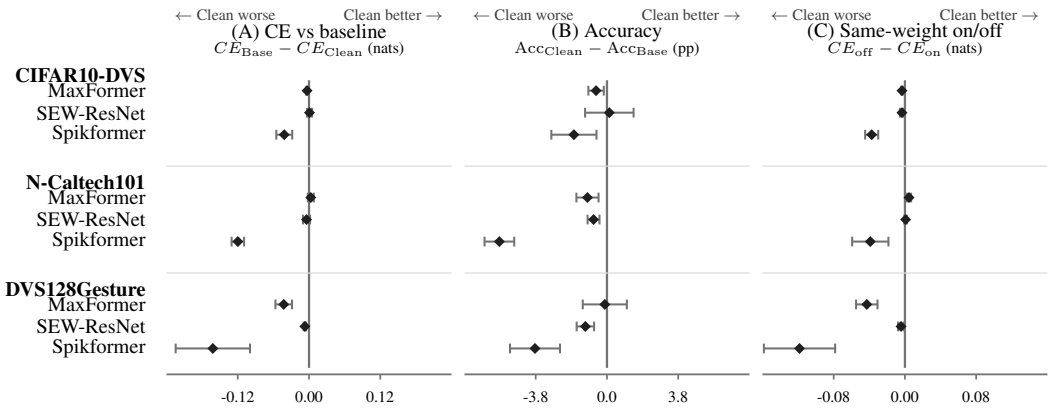


Figure 4: Natural residual effects depend on the evaluation endpoint. Positive values favor Clean. (A) Baseline-minus-Clean cross-entropy (CE), so a positive value means lower CE for Clean; (B) Clean-minus-Baseline accuracy in percentage points; and (C) same-weight off-minus-on CE, comparing the selected Clean model with its side path disabled and enabled. Each point is the mean of three network-level effects, after averaging eight bands and two adapter seeds within each network. Error bars denote descriptive standard errors across the three network-level effects; vertical lines mark zero. These development comparisons are not independent test estimates.

rate known-carrier check reads only the injection delta; it verifies supplied-signal readability, not sensitivity of the natural task path to missing content (Appendix F.3).

Natural residuals do not produce a consistent average improvement across the panel. MaxFormer/N-Caltech has a small mean CE reduction and positive on/off gain despite lower accuracy (Figure 4, A–C); SEW/CIFAR has a small accuracy increase but negative on/off CE gain. Spikformer has worse mean CE and accuracy on all three datasets under this short budget. Numerical endpoints are

in Table 7 in Appendix F.3. These descriptive comparisons neither rank architectures nor isolate a training-budget effect.

Beating a control also need not mean improving on Baseline. On CIFAR/MaxFormer, Phase-minus-Clean and Extra-Z-minus-Clean mean CE are $+0.0026$ and $+0.0005$ nats, yet Baseline-minus-Clean is -0.0034 nats. These descriptive means illustrate why Baseline and each control must be reported separately (Appendix Table 7). The original Phase residual is not energy-matched to Clean after reconstruction subtraction. In the CIFAR six-arm extension, its input energy materially differs in $80/144$ network–adapter–band cells; per-sample L2 matching removes the mismatch in all 144. Only 7 of the 80 mismatched cells meet the registered descriptive performance-shift flag, all in MaxFormer. Direct phase randomization of the Clean residual also gives architecture-dependent effects (Appendix Table 9). The Phase contrast therefore does not by itself identify a content-specific benefit.

Capacity and training budget offer another possible explanation for limited residual utility. At the same CIFAR stem boundary, a two-network sensitivity per architecture crossed 257-parameter versus full-rank adapters with 3 versus up to 30 head/adapter epochs for SEW and Spikformer. All capacity and budget Clean-CE contrast sample intervals include zero, and directions vary between networks. Ten versus one reconstructor epochs lower reconstruction MSE in both architectures, without a shared classification gain (Appendix Table 10). The longer budget allows early stopping and does not equate actual update counts. These checks do not produce a common explanation or a consistent benefit under the tested changes; other estimators and optimization schedules remain open.

5 Discussion and limitations

What changes in a frequency-based explanation? The experiments motivate a change in the unit of explanation, from the energy in one band to the response of a specified operator and predictor under intervention. The MaxFormer trace identifies where a source-band effect spreads; the fixed decoders measure how that perturbation changes their predictions. Original-head patches and natural residuals then test different forms of predictive benefit. These comparisons help formulate and audit restoration hypotheses without treating an off-diagonal response as evidence that task content has moved intact. Architectural gains based on frequency selection can remain useful even when their interpretation requires more than a low-pass argument.

What should a restoration claim predict? A useful next test is whether destination-decoder responses predict which patch targets benefit the original classifier under a fixed intervention budget. The present experiments motivate this question but do not validate such a prediction rule. A rule would need to be fixed before testing new networks or examples. Likewise, the negative binary joint result applies to one dose and selection procedure; it leaves other joint codes and interventions open (Williams & Beer, 2010). Natural restoration requires a further comparison against Baseline, content and capacity controls, and the same model with its side path disabled. An advantage over only one of these alternatives cannot identify missing upstream content.

Scope of the evidence. The experiments use short event-count sequences, so their spectral axis does not establish effects in physical time, long videos, or spatial filtering. Finite-window transients and phase-induced distribution shifts constrain operator and decoder interpretations. Fourier patches introduce nonbinary states, whereas binary patches sacrifice strict band isolation; their different joint responses do not isolate domain from dose. The natural panel has 200, 404, and 55 evaluation images per dataset. New-dataset evaluation images inform backbone checkpoint selection, and historical CIFAR exposure remains. Network-level SE excludes image-sampling and selection uncertainty. The residual sensitivity checks cover limited estimators and budgets, and the carrier check bypasses the natural task path. Independent, locked evaluation and stronger identifiability tests would be needed to attribute a predictive deficit to information erasure.

6 CONCLUSION

Frequency-based explanations require distinct measurements of operator response, decoder sensitivity, and task benefit. Our experiments locate cross-band perturbation effects at a MaxFormer's first LIF and record cross-band decoder responses across architectures. Same- and cross-band targets both select useful binary oracle patches, while the tested joint target does not recover performance at the same budget. Natural residuals show why a favorable control comparison must also be checked against Baseline. These development findings do not decide information preservation or loss. They make the practical requirement more specific: proposals to restore a frequency component should test which predictor benefits, under which intervention, and relative to which alternatives.

REPRODUCIBILITY STATEMENT

The accompanying implementation records operator boundaries, band definitions, seed roles, data partitions, and arm-specific results. The manuscript result figures and numerical tables are generated from saved artifacts with source hashes. Unified utility vectors were recomputed from paired per-image losses. Earlier patch CE and predictions were recomputed from saved logits; new joint, budget, and residual-control summaries use valid read-only audit exports. Repeated networks are not treated as new images. Appendix D describes the evaluation families and Appendix A provides derivations. Numerical checks validate the stated finite-dimensional identities and counterexamples; they are not substitutes for natural-data experiments. No additional model training or confirmatory test evaluation was performed when preparing this draft.

AI USE STATEMENT

Generative AI tools assisted with research discussion, formulation and checking of mathematical arguments, experimental and software design, interpretation of saved results, literature searches, manuscript drafting, and preparation of figures and references. Result figures are plotted from recorded experimental artifacts. Figure 1 is a conceptual illustration supplied by the authors, not an experimental observation.

REFERENCES

Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza, Jeff Kusnitz, Michael Debole, Steve Esser, Tobi Delbruck, Myron Flickner, and Dharmendra Modha. A low power, fully event-based gesture recognition system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. URL https://openaccess.thecvf.com/content_cvpr_2017/papers/Amir_A_Low_Power_CVPR_2017_paper.pdf.

Nora Belrose, David Schneider-Joseph, Shauli Ravfogel, Ryan Cotterell, Edward Raff, and Stella Biderman. LEACE: Perfect linear concept erasure in closed form. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/d066d21c619d0a78c5b557fa3291a8f4-Abstract-Conference.html.

Nicolas Brunel, Frances S. Chance, Nicolas Fourcaud, and L. F. Abbott. Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Physical Review Letters*, 86(10): 2186–2189, 2001. doi: 10.1103/PhysRevLett.86.2186. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.86.2186.

Wei Fang, Zhaofei Yu, Yanqi Chen, Tiejun Huang, Timothée Masquelier, and Yonghong Tian. Deep residual learning in spiking neural networks. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2102.04159.

Yuetong Fang, Deming Zhou, Ziqing Wang, Hongwei Ren, Zecui Zeng, Lusong Li, Shibo Zhou, and Renjing Xu. Spiking neural networks need high-frequency information. In *Advances in Neural Information Processing Systems*,

2025. URL https://papers.nips.cc/paper_files/paper/2025/file/956834836f36dd07df7064ff42ca69f2-Paper-Conference.pdf.

Shibo Feng, Wanjin Feng, Xingyu Gao, Peilin Zhao, and Zhiqi Shen. TS-LIF: A temporal segment spiking neuron network for time series forecasting. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/45bd303ea32a8193d3c19de5d5eaf138-Abstract-Conference.html.

Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Joerg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2022. doi: 10.1109/TPAMI.2020.3008413. URL https://arxiv.org/abs/1904.08405.

Atticus Geiger, Hanson Lu, Thomas Icard, and Christopher Potts. Causal abstractions of neural networks. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2106.02997.

John Hewitt and Percy Liang. Designing and interpreting probes with control tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 2733–2743, 2019. doi: 10.18653/v1/D19-1275. URL https://aclanthology.org/D19-1275/.

John Hewitt, Kawin Ethayarajh, Percy Liang, and Christopher Manning. Conditional probing: measuring usable information beyond a baseline. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1626–1639, 2021. doi: 10.18653/v1/2021.emnlp-main.122. URL https://aclanthology.org/2021.emnlp-main.122/.

Saya Higuchi, Sebastian Kairat, Sander Bohte, and Sebastian Otte. Balanced resonate-and-fire neurons. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 18305–18323, 2024. URL https://proceedings.mlr.press/v235/higuchi24a.html.

Thomas E. Huber, Jules Lecomte, Borislav Polovnikov, and Axel von Arnim. Scaling up resonate-and-fire networks for fast deep learning. In *Computer Vision – ECCV 2024 Workshops*, pp. 241–258, 2025. doi: 10.1007/978-3-031-92460-6_15. URL https://doi.org/10.1007/978-3-031-92460-6_15.

Eugene M. Izhikevich. Resonate-and-fire neurons. *Neural Networks*, 14(6–7):883–894, 2001. doi: 10.1016/S0893-6080(01)00078-8. URL https://www.izhikevich.org/publications/resfire.htm.

Hongmin Li, Hanchao Liu, Xiangyang Ji, Guoqi Li, and Luping Shi. CIFAR10-DVS: An event-stream dataset for object classification. *Frontiers in Neuroscience*, 11:309, 2017. doi: 10.3389/fnins.2017.00309. URL https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2017.00309/full.

Rakesh Malladi, Don H. Johnson, Giridhar P. Kalamangalam, Nitin Tandon, and Behnaam Aazhang. Mutual information in frequency and its application to measure cross-frequency coupling in epilepsy. *IEEE Transactions on Signal Processing*, 66(11):3008–3023, 2018. doi: 10.1109/TSP.2018.2821627. URL https://arxiv.org/abs/1711.01629.

Aaron Mueller, Atticus Geiger, Sarah Wiegreffe, Dana Arad, Iván Arcuschin, Adam Belfki, Yik Siu Chan, Jaden Fried Fiotto-Kaufman, Tal Haklay, Michael Hanna, Jing Huang, Rohan Gupta, Yaniv Nikankin, Hadas Orgad, Nikhil Prakash, Anja Reusch, Aruna Sankaranarayanan, Shun Shao, Alessandro Stolfo, Martin Tutek, Amir Zur, David Bau, and Yonatan Belinkov. MIB: A mechanistic interpretability benchmark. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 45069–45108, 2025. URL https://proceedings.mlr.press/v267/mueller25a.html.

Alan V. Oppenheim and Jae S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69 (5):529–541, 1981. doi: 10.1109/PROC.1981.12022. URL https://ieeexplore.ieee.org/document/1456290/.

Garrick Orchard, Ajinkya Jayawant, Gregory Cohen, and Nitish Thakor. Converting static image datasets to spiking neuromorphic datasets using saccades. *Frontiers in Neuroscience*, 9:437, 2015. doi: 10.3389/fnins.2015.00437. URL https://arxiv.org/abs/1507.07629.

Namuk Park and Songkuk Kim. How do vision transformers work? In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=D78Go4hVcxO.

Tiago Pimentel, Josef Valvoda, Rowan Hall Maudslay, Ran Zmigrod, Adina Williams, and Ryan Cotterell. Information-theoretic probing for linguistic structure. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4609–4622, 2020. doi: 10.18653/v1/2020.acl-main.420. URL https://aclanthology.org/2020.acl-main.420/.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5301–5310, 2019. URL https://proceedings.mlr.press/v97/rahaman19a.html.

James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1–4):77–94, 1992. doi: 10.1016/0167-2789(92)90102-S. URL https://doi.org/10.1016/0167-2789(92)90102-S.

Elena Voita and Ivan Titov. Information-theoretic probing with minimum description length. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 183–196, 2020. doi: 10.18653/v1/2020.emnlp-main.14. URL https://aclanthology.org/2020.emnlp-main.14/.

John J. Wade, Liam J. McDaid, Jose A. Santos, and Heather M. Sayers. SWAT: A spiking neural network training algorithm for classification problems. *IEEE Transactions on Neural Networks*, 21(11):1817–1830, 2010. doi: 10.1109/TNN.2010.2074212. URL https://doi.org/10.1109/TNN.2010.2074212.

Kevin Ro Wang, Alexandre Variengien, Arthur Conmy, Buck Shlegeris, and Jacob Steinhardt. Interpretability in the wild: a circuit for indirect object identification in GPT-2 small. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2211.00593.

Peihao Wang, Wenqing Zheng, Tianlong Chen, and Zhangyang Wang. Anti-oversmoothing in deep vision transformers via the Fourier domain analysis: From theory to practice. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2203.05962.

Paul L. Williams and Randall D. Beer. Nonnegative decomposition of multivariate information, 2010. URL https://arxiv.org/abs/1004.2515.

Wenjie Wu, Dexuan Huo, and Hong Chen. SpikF: Spiking Fourier network for efficient long-term prediction. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 67342–67368, 2025. URL https://proceedings.mlr.press/v267/wu25m.html.

Yilun Xu, Shengjia Zhao, Jiaming Song, Russell Stewart, and Stefano Ermon. A theory of usable information under computational constraints. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/2002.10689.

Shuhan Ye, Yuanbin Qian, Yi Yu, Chong Wang, Yuqi Xie, Jiazhen Xu, Kun Wang, and Xudong Jiang. Fire on motion: Optimizing video pass-bands for efficient spiking action recognition, 2026. URL https://arxiv.org/abs/2601.22675.

Dong Yin, Raphael Gontijo Lopes, Jon Shlens, Ekin Dogus Cubuk, and Justin Gilmer. A Fourier perspective on model robustness in computer vision. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/b05b57f6add810d3b7490866d74c0053-Abstract.html.

Kairong Yu, Tianqing Zhang, Hongwei Wang, and Qi Xu. FSTA-SNN: Frequency-based spatial-temporal attention module for spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22227–22235, 2025. doi: 10.1609/aaai.v39i21.34377. URL https://ojs.aaai.org/index.php/AAAI/article/view/34377.

Fred Zhang and Neel Nanda. Towards best practices of activation patching in language models: Metrics and methods. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Hf17y6u9BC.

Zhaokun Zhou, Yuesheng Zhu, Chao He, Yaowei Wang, Shuicheng Yan, Yonghong Tian, and Li Yuan. Spikformer: When spiking neural network meets transformer. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=frE4fUwz_h.

A Fourier statements and their scope

A.1 Frequency preservation

Let E_k^{in} and E_k^{out} denote input/output subspaces supported only at DFT bin k, and let S denote one-step circular delay. For a fixed linear map A, the statements

$$AE_k^{\text{in}} \subseteq E_k^{\text{out}} \quad \text{for all } k, \qquad \widetilde{A}_{\ell k} = 0 \quad (\ell \neq k), \qquad S_{\text{out}} A = A S_{\text{in}}$$

are equivalent. To see this, diagonalize the circular shift by the unitary DFT. Its eigenvalues $\rho_k = \exp(-2\pi ik/T)$ are distinct, and the intertwining identity becomes $(\rho_\ell - \rho_k)\widetilde{A}_{\ell k} = 0$. This forces every off-diagonal frequency block to vanish. Conversely, block diagonality implies the intertwining identity. In the time domain, the same condition states that the channel-valued kernel depends only on the time difference modulo T, yielding circular convolution.

A shared per-step spatial map is the special case $(Ax)_t = Mx_t$. Its channel/spatial matrix M need not be invertible. A time-varying linear map can mix frequencies even though it is linear; therefore linearity alone is not a sufficient control. Constant temporal bias occupies only DC. Evaluation BN has fixed coefficients, whereas training BN uses input-dependent statistics, invalidating this fixed-map argument for general perturbation comparisons.

A.2 Finite-window integration

For $u_t = \alpha u_{t-1} + \beta x_t$, $0 < \alpha < 1$, the steady-state response is

$$|H(\omega)| = \frac{|\beta|}{\sqrt{1 + \alpha^2 - 2\alpha\cos\omega}},$$

which decreases on $[0, \pi]$ for $\beta \neq 0$. In a finite window,

$$\widehat{u}[k] = \frac{\beta \widehat{x}[k]}{1 - \alpha e^{-i\omega_k}} + \frac{\alpha(u_{-1} - u_{T-1})}{1 - \alpha e^{-i\omega_k}}, \qquad \omega_k = 2\pi k/T. \tag{5}$$

Use the nonunitary DFT $\widehat{x}[k] = \sum_{t=0}^{T-1} x_t z_k^t$ with $z_k = e^{-2\pi ik/T}$. The recurrence $u_t = \alpha u_{t-1} + \beta x_t$ gives

$$\sum_{t=0}^{T-1} u_{t-1} z_k^t = u_{-1} + z_k \widehat{u}[k] - u_{T-1}.$$

Here $z_k^T = 1$; substituting yields Eq. (5). For a pure grid tone $x_t = ae^{i\omega_k t}$ and zero initial state,

$$u_t = H_k a e^{i\omega_k t} - H_k a e^{-i\omega_k} \alpha^{t+1}, \qquad H_k = \frac{\beta}{1 - \alpha e^{-i\omega_k}}.$$

The second term generally has broad finite-DFT support. Matching the periodic initial state removes it; finite burn-in only reduces its magnitude. In particular, observing off-diagonal finite-window energy after a leaky recurrence does not by itself prove threshold mixing.

A.3 Mixing, threshold, and reset

Expanding the inverse DFTs of x_t and y_t and using discrete orthogonality gives

$$\widehat{xy}[k] = \frac{1}{T} \sum_j \widehat{x}[j] \widehat{y}[k - j].$$

For per-time attention contraction, the corresponding identity is

$$\widehat{QK^\top V}[k] = \frac{1}{T^2} \sum_{a+b+c=k \pmod T} \widehat{Q}[a] \widehat{K}[b]^\top \widehat{V}[c].$$

Matrix transposition here acts on the token/feature indices, not on the time axis. Fixed scalar attention normalization does not change the frequency-support argument. This expression describes possible mixing, not the magnitude or usefulness of individual terms in a trained network.

Thresholded periodic signals need not retain the driving waveform's Fourier support. For example, thresholding a cosine above a nonzero level produces a pulse whose Fourier series contains multiple harmonics. Their presence and magnitude depend on the threshold and waveform; not every harmonic is nonzero for every threshold. Reset changes future membrane states through the generated spikes. Within a fixed spike pattern, the membrane recurrence has a simpler piecewise structure, but crossing a threshold changes the pattern discontinuously. Surrogate derivatives used for training should not be interpreted as the ordinary derivative of this hard forward map.

For a twice differentiable approximation F, local Fourier coordinates give a first-order term $\widetilde{J\,\delta x}$ and a second-order bilinear term. Off-diagonal Jacobian blocks describe local cross-frequency response, and the Hessian describes pairwise mixing. These are local statements with transform-dependent normalizations, not a globally differentiable account of hard spikes.

B Probe risk and operational rescue utility

For predictor family $\mathcal{G}$, define $R^*_{\mathcal{G}}(S) = \inf_{g\in\mathcal{G}_S} \mathbb{E}[-\log g(Y \mid S)]$ and $A_{\mathcal{G}}(S) = R^*_{\mathcal{G}}(S) - H(Y \mid S)$. The optimal family-risk contrast obeys

$$R^*_{\mathcal{G}}(Z) - R^*_{\mathcal{G}}(Z, A) = I(Y; A \mid Z) + A_{\mathcal{G}}(Z) - A_{\mathcal{G}}(Z, A). \tag{6}$$

For any predictor g and input S, log-loss decomposes as

$$R(g; S) = H(Y \mid S) + \mathbb{E}_S D_{\mathrm{KL}}(p(\cdot \mid S) \,\|\, g(\cdot \mid S)).$$

Taking the infimum over the predictor family identifies $A_{\mathcal{G}}(S)$ as the smallest achievable expected conditional KL. Subtracting the Z and (Z, A) equations proves Eq. (6). If the augmented family can ignore A, the optimum risk contrast is nonnegative. Its relation to conditional MI nevertheless depends on the difference between the two approximation gaps.

For trained probes $\widehat{g}_Z, \widehat{g}_{ZA}$, define $e_S = R(\widehat{g}_S; S) - R^*_{\mathcal{G}}(S)$. Their population contrast is

$$R(\widehat{g}_Z; Z) - R(\widehat{g}_{ZA}; Z, A) = I(Y; A \mid Z) + A_{\mathcal{G}}(Z) - A_{\mathcal{G}}(Z, A) + e_Z - e_{ZA}.$$

Finite held-out averages introduce an additional estimation discrepancy. This prevents identifying a trained probe's CE difference with Shannon information simply because the loss is cross-entropy.

The historical secondary statistic compares separately trained arms:

$$\overline{V}_G = \frac{1}{n}\sum_i [\min(L_{P,i}, L_{E,i}) - L_{C,i}]. \tag{7}$$

It obeys

$$\overline{V}_G \le \min(\overline{L}_P, \overline{L}_E) - \overline{L}_C.$$

For illustration, let two samples have Phase losses $(0, 2)$, Extra-Z losses $(2, 0)$, and Clean losses $(0.5, 0.5)$. Clean improves on each control's mean by 0.5, but $\overline{V}_G = -0.5$. Thus a negative canonical value need not mean Clean loses to either control in mean CE. Since Baseline cancels algebraically, a positive V_G alone does not establish improvement on Baseline. Writing $d_i = L_{P,i} - L_{E,i}$ makes the control-disagreement term explicit:

$$\overline{V}_G = \min(\overline{L}_P, \overline{L}_E) - \overline{L}_C - \tfrac{1}{2}\left(\overline{|d|} - |\overline{d}|\right).$$

The final term is nonnegative and reflects disagreement between the controls, not an information-loss quantity. It vanishes when one control has no greater loss than the other on every sample. Neither sign of this statistic identifies erasure, and a negative sign alone does not exclude predictive benefits under other stated metrics.

For an independent evaluation set and fixed predictors with clipped per-sample contrast $\Delta_i \in [-B, B]$, Hoeffding's inequality and a union bound give a deviation at most $B\sqrt{2\log(2J/\delta)/n}$ simultaneously over J fixed contrasts, with probability at least $1 - \delta$. This statement does not directly apply to the unbounded, selection-used historical CE assays. Their bootstrap proportions are operational summaries, not a remedy for adaptive selection or a proof of nominal family-wise coverage.

Erasure, accessibility, and a side path. For upstream U, downstream $Z = F(U)$, and a deterministic downstream-only transformation $H(Z)$, the data-processing inequality gives $I(Y; H(Z)) \leq I(Y; Z)$. A downstream transformation can still improve a constrained classifier by making existing dependencies easier to read. An upstream-access side path A can add information only to the extent $I(Y; A \mid Z) > 0$; a trained-system improvement alone does not identify that quantity.

Marginal frequency tests also miss joint codes. For independent bits A_1, A_2 and $Y = A_1 \text{ XOR } A_2$, each single input has zero mutual information with Y, while the pair determines Y. Redundant bands give the converse warning against adding marginal gains. Averages of losses from separately trained band adapters are not joint-band interventions and cannot measure synergy.

C Controlled carrier calibration

Let $q \in \{0, 1\}$ be the auxiliary label, $S = (-1)^q$, and u a fixed unit-RMS carrier with a channel code. If s_{pre} denotes the fitting-split upstream RMS, the constructed states are

$$Z_{\text{no}} = Z_{\text{pre}} + \lambda s_{\text{pre}} S u, \quad Z_{\text{erase}} = Z_{\text{pre}} + (1 - \rho)\lambda s_{\text{pre}} S u, \quad \Delta Z_q = \rho \lambda s_{\text{pre}} S u.$$

These are states at the pre-spiking boundary, not raw sensor inputs. Thus $Z_{\text{no}} - Z_{\text{erase}} = \Delta Z_q$ is an engineering identity whose floating-point error is checked. The label q is not a quality parameter. Balancing it within natural classes controls class imbalance, but does not prove population independence from all incidental input features. Carrier-absent and held-out readouts remain necessary controls. Natural Y is retained only for a performance-drift check; it is not the q target or a rescue gate.

An idealized Gaussian template channel illustrates the interpretation. If $W = aS + \sigma Z$ with $Z \sim \mathcal{N}(0, 1)$ independent of the balanced bit, Bayes accuracy is $\Phi(a/\sigma)$ and Bayes log-loss is

$$\mathcal{L}(r) = \mathbb{E}_{Z \sim \mathcal{N}(0,1)} \log(1 + \exp[-2r(r + Z)]), \qquad r = a/\sigma.$$

For fixed $0 < \rho < 1$, the erasure gap $\mathcal{L}((1-\rho)r) - \mathcal{L}(r)$ tends to zero both as $r \to 0$ and as $r \to \infty$. At large amplitude the residual carrier remains highly readable. Increasing baseline amplitude at fixed erasure fraction therefore need not increase the predictive-loss gap monotonically. For fixed positive r, however, the gap increases with ρ.

This channel is not asserted to describe real post-LIF features. Hard thresholds can make sufficiently small perturbations invisible and cause abrupt changes at a background-dependent margin. The smallest successful grid point in the actual experiment depends on optimization, precision, sample size, and readout, and cannot be promoted to a universal detection lower bound.

The ideal channel models a carrier observed together with background noise. The empirical cross-architecture calibration instead reads only the target-band injection delta, excluding background activations and bypassing downstream spiking computation. It checks supplied-signal readout, not finite-SNR recovery through the natural task path. Its conditions appear in Appendix F.1. Historical graded-dose results are reported separately in Appendix H.

D Experimental protocols and exposure

Preprocessing and frequency resolution. Event frames retain polarity and use event-count integration. N-Caltech101 is padded to a square and resized to 128×128. Because event counts define the temporal bins, their physical durations can differ. The main results characterize the resulting sequence axis. With $T = 16$, b_0 contains DC and b_7 includes both the last non-Nyquist bin and Nyquist; the latter has no arbitrary complex phase for real signals. Spectral power in the original MaxFormer implementation uses the nonunitary rFFT, doubles conjugate-paired interior-bin powers, and leaves DC and Nyquist weights at one. Power is averaged over non-time feature coordinates; samples and phase seeds are accumulated before row normalization. The equivalent time-domain squared norm differs by the fixed DFT normalization factor. This factor cancels in the reported within-row shares. Negative DC source results under an identity phase operation cannot establish that the model ignores static information.

Table 2: Evaluation families are not interchangeable. Numbers describe fitting/evaluation samples, not universal dataset splits.

Family	Samples	Interpretation
MaxFormer lineage, CIFAR	9000/1000	Historical test-based mechanism exploration
MaxFormer lineage, N-Caltech	7795/914	Historical floor partition; not later ceil results
MaxFormer lineage, Gesture	1176/288	Historical gesture train/test analysis
SEW all-band lineage	9000/1000	Historical test analysis, distinct from validation-selected single-band pilot
Spikformer lineage	800/200	Fixed-checkpoint screening on training-side data
Historical natural CIFAR full-sample	8100/900	Validation also used for selection; development only
CIFAR q / natural lineage / utility	800/200/200	Shared probe roles; historical backbone exposure; development evidence
Natural N-Caltech follow-up	7078/808	Dataset-specific train-side validation
Natural Gesture follow-up	1066/110	Dataset-specific train-side validation
Function-preserving recovery	6480/810/810	Disjoint fit/select/score within training core; source backbone trained on that core

Trace and state controls. The MaxFormer and SEW traced paths reproduce untraced outputs with recorded maximum absolute error zero. Spikformer likewise records zero trace-on/off logit discrepancy. Non-DC source perturbations in the MaxFormer input matrices remain on the diagonal to numerical precision. Historical DC normalization can amplify a near-zero numerical residual; we mask that source row rather than plot it as a valid experiment. Each SNN forward is followed by a neuron-state reset. Evaluation BN statistics are fixed in frozen representation probes.

Readout and seed roles. The original MaxFormer lineage uses spatial pooling and CountSketch features of dimension 128 with ridge parameter 1, repeated over three projection and three phase seeds. These repeats assess probe/intervention variability at one checkpoint. Network training seeds, projection seeds, phase seeds, and adapter seeds are reported separately. The 30-epoch Spikformer follow-up freezes the backbone and fits a fresh 2570-parameter head with a 257-parameter adapter. Its four cells are network seeds 35308/35309 crossed with adapter seeds 9101/9102. The same validation set selects arm checkpoints and supplies reported paired summaries, so no claim of independent confirmation follows.

Unified natural lineage implementation checks. The source replacement is $X - P_b X + \mathrm{PhaseShuffle}_b(X)$. The prior isolated-band implementation also removed non-target frequencies and is excluded from the current propagation results. All nine v2 tasks passed a spectral preservation check, with maximum non-target discrepancy 1.34×10^{-7} relative to the input spectrum. The fixed readout uses AdamW (learning rate 10^{-3}, weight decay 10^{-4}), minibatch 64, at most 30 epochs, and patience 5. The selected states are frozen. Both the spatial representation and optimization budget changed from the excluded implementation, so their individual effects are not isolated.

The unified auxiliary energy diagnostic is the absolute change in normalized band power shares; it is not the perturbation power $Q_{\ell,bc}$ of Equation 2. The main unified routing figure uses decoder CE changes. Historical perturbation-power figures retain their original estimator. DC sign controls and AC phase controls are identified separately; Nyquist has no arbitrary complex phase and is left unchanged by the unified phase intervention, including within b_7.

CIFAR network exposure and training. MaxFormer and Spikformer use network seeds 35307, 35308, and 35309. SEW uses the historical seed 2020 and added seeds 35308/35309. The historical SEW checkpoint trained on 9,000 images and used test performance in model selection; it cannot supply an independent evaluation on the current training-side manifest. The added SEW networks use the frozen 64-epoch SGD recipe and validation selection without new test evaluation. New probe fit/select/evaluation assignments do not remove any prior exposure. Utility uses adapter seeds 9101/9102, three epochs with patience two, AdamW (learning rate 10^{-3}, weight decay 10^{-4}), and one epoch of reconstruction fitting. SEW uses batch 8 after the registered batch-16 memory fail-

Table 3: Full eight-operator extrapolation MSE, mean ± sample SD. Mechanism columns separate integration, pointwise nonlinearity, and reset. – denotes an absent component. This is the new matched matrix; historical four-operator numbers are not pooled with it. Identity errors below the shown precision round to zero.

Operator	Temporal integration	Pointwise nonlinearity	Reset	Identity	Doubling	Difference	Nuisance
No integration							
Identity	–	Linear	–	$\mathbf{0.0000 \pm 0.0000}$	0.5054 ± 0.0011	0.5126 ± 0.0061	0.4051 ± 0.0016
ReLU	–	ReLU	–	0.0000 ± 0.0000	0.5054 ± 0.0011	$\mathbf{0.3037 \pm 0.0037}$	$\mathbf{0.3519 \pm 0.0017}$
tanh	–	tanh	–	0.0001 ± 0.0000	0.5054 ± 0.0011	0.3052 ± 0.0040	0.3521 ± 0.0013
IIR, no reset							
Linear-IIR	IIR	Linear	–	0.1528 ± 0.0003	0.5054 ± 0.0011	0.5126 ± 0.0061	0.3768 ± 0.0015
IIR+ReLU	IIR	ReLU	–	0.1479 ± 0.0033	0.3002 ± 0.0317	0.3490 ± 0.0041	0.3784 ± 0.0019
IIR+tanh	IIR	tanh	–	0.1496 ± 0.0033	0.2956 ± 0.0060	0.3477 ± 0.0044	0.3780 ± 0.0016
IIR+Threshold	IIR	Threshold	–	0.1462 ± 0.0007	$\mathbf{0.2833 \pm 0.0040}$	0.3503 ± 0.0045	0.3790 ± 0.0018
IIR + threshold + reset							
LIF	IIR	Threshold	Subtractive	0.0118 ± 0.0004	0.3895 ± 0.0434	0.3531 ± 0.0039	0.3631 ± 0.0028

ure; the other architectures use 16. Without gradient accumulation this also changes the number of updates per epoch, a remaining budget difference. A 257-parameter adapter is a full map for MaxFormer and rank-2/rank-4 maps for SEW/Spikformer; equal parameter count does not imply identical function classes. Reconstructor and head sizes also depend on channel count. The extra scalar parameter is the zero-initialized learned injection scale η; all channel maps are bias-free. Thus zero-content Baseline has no additive offset. Incomplete reconstruction can leave downstream-redundant content in R_b, and joint head/adapter fitting may be limited by the short budget. Neither explanation is isolated by the current comparison.

Legacy statistics and Phase behavior. Earlier MaxFormer/SEW result files use a legacy scalar V equal to the minimum of separate mean Baseline and control contrasts. Some historical files do not contain the later canonical V_G at all. Spikformer files use the sample-level statistic in Eq. (7). The manuscript therefore does not pool their numeric utility values. The historical Phase arm uses step-dependent randomness when sample IDs are unavailable; later retained-head/matched-fine-tuning paths fix intervention randomness by sample ID. Neither setting alone guarantees that subtracting the reconstructor preserves residual-spectrum matching.

Selection and checkpoints. Train/validation ID disjointness is necessary but insufficient for independence from the entire learning procedure. A score set drawn from the backbone's original training core is new to an adapter, not new to the pretrained backbone. Checkpoint selection on historical test performance and repeated use of validation data are separately acknowledged. A test-not-loaded field describes that run's behavior, not the complete history of every sample. We retain historical provisional flags without interpreting them as post-selection statistical guarantees.

E CONTROLLED OPERATOR DETAILS AND NUMERICAL RESULTS

The eight operators share the affine architecture and training budget, not learned weights. With current i_t, integrating conditions use $u_t^- = 0.5u_{t-1} + i_t$. Linear-IIR outputs u_t^-; IIR+ReLU/tanh apply the corresponding activation and retain $u_t = u_t^-$. IIR+Threshold outputs $\mathbf{1}[u_t^- \geq 1]$ without reset. LIF outputs the same threshold rule and sets $u_t = u_t^- - s_t$. Static conditions apply Identity/ReLU/tanh directly to i_t. The fixed surrogate slope is 5 for threshold training.

Training frequencies occupy $[0.035, 0.075] \cup [0.115, 0.155]$ cycles per step; interpolation occupies $[0.080, 0.110]$ and extrapolation $[0.165, 0.205]$. Each of three training repeats uses 80 epochs with ten steps each, batch 256, learning rate 0.003, weight decay 0.0001 and gradient clipping 1.0. Evaluation uses 1,024 generated sequences per condition. The formal matrix has 96 task–operator–repeat records. Difference-frequency targets retain the amplitude product and phase difference; nuisance tasks retain the low-frequency signal while adding a stronger high-frequency input.

Static ReLU/tanh produce sum/difference components after their nonlinear outputs. The recorded affine biases mean the tanh traces are not the zero-bias symmetric case with only odd harmonics. A lack of useful doubling in the fitted static models is not proof of representational impossibility. The

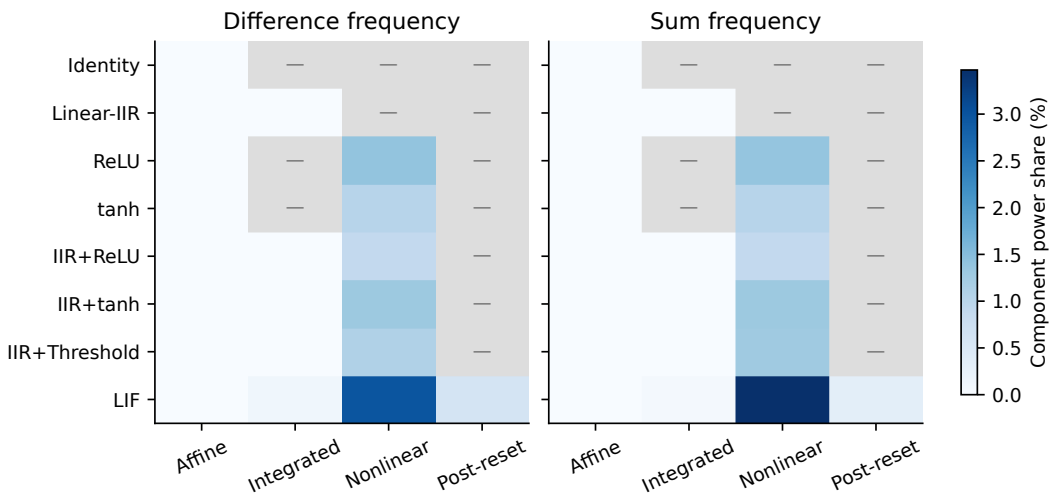


Figure 5: Fixed-input response traces from one preselected doubling-trained state per operator. Two tones at 0.046875 and 0.15625 cycles/step are sampled over 512 steps, discarding the first 128. The plot shows fitted difference- and sum-frequency power shares at available operator boundaries; gray cells mark absent stages. The shared color scale is in percent. LIF integrated states already include feedback from earlier resets. It is an illustrative response audit, not an estimate of between-training uncertainty. Frontend weights and biases differ between fitted operators; residual power also includes unmodeled components and numerical/window effects.

Table 4: Direct dual-tone spectral response of selected fitted operators. Difference and sum components are at 0.109375 and 0.203125 cycles/step; entries are percent of output power. One preselected doubling-trained state per operator receives the same fixed input, discarding the first 128 of 512 steps. The linear value rounds to zero. These illustrative spectra do not estimate variation across training repeats or match learned affine weights across operators.

Operator	Difference	Sum
Identity (linear)	0.00	0.00
ReLU	1.39	1.38
tanh	1.03	1.03
LIF	2.98	3.47

synthetic LIF has subtractive reset; natural networks include hard-reset nodes. Capability results from one reset rule do not establish the dynamics or task benefit of every natural spiking node.

F Calibration and natural-panel details

F.1 Known-content calibration: CIFAR reference

At each erasure fraction, all rescue arms share $Z_{\mathrm{pre}} + (1-\rho)C_q$. Clean supplies ρC_q, Phase supplies a q-independent phase control with matched RMS, Wrong-band supplies a q-coded b_2 carrier, and Extra-Z supplies a signed permutation of the erased downstream state. The q head reads only the target-band adapter delta. This makes Wrong-band rejection and zero-delta Erase/off reference behavior properties of the readout design. No-Erase uses a separate complete-boundary readout; comparing its CE directly with the delta-only Clean head is not a measure of the fraction of original information restored. The reference q head trains for eight full-fit steps; rescue q heads and adapters train for three epochs, including their initial state in selection.

No-Erase CE is $0.67348/0.56088/0.66506$ for MaxFormer/SEW/Spikformer, at 100% accuracy in each. Its weak confidence explains why dividing by the No-Erase–Erase CE gap can yield a large recovery ratio; we use absolute CE and same-weight gains in the main text. At partial erasure, Extra-

Z can decode the surviving q carrier, and should not be interpreted as a failed negative control. At zero erasure, the removed carrier and Clean delta vanish. Natural-label performance is a separate drift monitor.

F.2 CIFAR READOUTS AND ARM MEANS

Table 5 reports the CIFAR readout summaries. Full-panel arm means and diagnostic curves appear in Table 7 and Figure 7.

Table 5: Unified natural readouts. Accuracy is in percent. Single-band ranges are over eight band means, each averaged over three networks. Joint split concatenates eight 128-D sketches (1,024 inputs); single-band heads and the summed joint representation each use 128 inputs. Routing entries are mean CE increases per cell, including the separate DC sign control. They are not row-normalized information shares.

Architecture	Single-band acc	Joint acc	Diagonal CE	Off-diagonal CE
MaxFormer	20.3–29.3	35.5	0.0845	0.00695
SEW-ResNet	16.8–32.0	38.0	0.2020	0.01933
Spikformer	19.8–32.7	34.8	0.1743	0.01721

F.3 REPAIRED THREE-ARCHITECTURE, THREE-DATASET PANEL

The expanded panel retains the first native continuous-to-spike boundary, $T = 16$, $K = 8$, two adapter initializations per network, and the same lineage/readout recipes. N-Caltech uses 101 classes, a $7078/808$ backbone train/validation split, and $1212/404/404$ fit/select/evaluation images. Gesture uses 11 classes, a $1066/110$ backbone split and $847/55/55$ assay images. CIFAR retains the earlier protocol described above. The head size depends on the number of classes; equal adapter parameter counts do not imply identical function classes across architectures.

The initially reused SEW and Spikformer checkpoints on the two new datasets had gradient-training exposure to evaluation images. They are excluded from the current panel. Four replacement backbones were trained under the frozen manifests, followed by all dependent health, lineage, natural, and q assays. A fresh audit checks exact train/validation ID membership, zero evaluation–train overlap, and matching checkpoint hashes across the downstream artifacts. All 27 health checks pass. The 18 new-dataset networks nevertheless use the evaluation images for backbone validation selection. Thus the repaired panel remains development-only; it is not three independently held-out replications per cell.

Table 7 gives arm means and paired contrasts alongside Figure 4. Means first combine eight bands and two adapters within each network, then combine three networks. Band-specific curves in Figure 7 instead average adapters within each band and network. Their negative values do not establish loss or preservation. The uncertainty summarizes network variation conditional on the fixed images and protocol; it does not include all selection or sampling uncertainty.

All nine fixed-erased-state checks are exact in the saved records. At complete erasure, control CE is close to $\log 2$ throughout the panel. The Phase arm on Spikformer/Gesture reaches $34/55 = 61.8\%$ accuracy; this finite-sample result should not be labeled exactly chance or taken as evidence of reliable content recovery. Gesture's odd evaluation count also prevents an exactly balanced binary label set. No additional q repetitions or official-test evaluation are implied by this table.

Residual-content and energy controls at the CIFAR stem. The six-arm extension repeats every three-architecture, three-network, two-adapter, eight-band CIFAR cell on the registered $800/200/200$ fit/select/evaluation IDs. Original Phase uses $R_P = \mathrm{PhaseShuffle}_b(Z_{\mathrm{pre}}) - f_b(Z_{\mathrm{post}})$; Clean uses $R_C = P_b Z_{\mathrm{pre}} - f_b(Z_{\mathrm{post}})$. Phase L2 matching rescales the original sample-ID Phase residual to $\|R_C\|_2$ per sample. The other new control phase-randomizes R_C directly, preserving its input magnitude spectrum. All six arms share head/adapter initialization, sample order, optimization budget, and frozen backbone.

Original Phase adapter-input energy is materially mismatched in $80/144$ network–adapter–band cells under the registered 1%-of-Clean rule. The scaled Phase control matches in $144/144$. Only

Table 6: Unified 3×3 first-stem decoder response (CE nats). Three network matrices are averaged within each architecture–dataset cell before summarizing AC source bands b_1–b_7. Diagonal and off-diagonal columns are cell means; the range gives the minimum and maximum off-diagonal cell responses. DC source sign controls are excluded; destination DC is included. Positive means worse fixed-decoder CE after intervention. These descriptive responses do not estimate transported information.

Dataset	Architecture	Diagonal	Off-diagonal	Off range
CIFAR10-DVS	MaxFormer	+0.059	+0.006	[-0.010, +0.040]
CIFAR10-DVS	SEW-ResNet	+0.080	+0.018	[-0.002, +0.078]
CIFAR10-DVS	Spikformer	+0.059	+0.015	[-0.004, +0.054]
N-Caltech101	MaxFormer	+0.029	+0.023	[+0.008, +0.079]
N-Caltech101	SEW-ResNet	+0.041	+0.015	[+0.000, +0.088]
N-Caltech101	Spikformer	+0.015	+0.010	[+0.000, +0.030]
DVS128Gesture	MaxFormer	+0.024	+0.016	[+0.001, +0.035]
DVS128Gesture	SEW-ResNet	-0.007	-0.004	[-0.031, +0.016]
DVS128Gesture	Spikformer	+0.007	-0.001	[-0.017, +0.013]

Table 7: Repaired 3×3 natural-task outcomes. Panel A reports mean CE (nats)/accuracy (percent) for each arm, averaging eight bands and two adapter seeds within each network before averaging three networks. Panel B reports paired effects as mean ± descriptive SE across three network means. Positive effects favor Clean. On/off CE is the same-weight $\mathrm{CE}_{\mathrm{off}} - \mathrm{CE}_{\mathrm{on}}$ contrast. These development results are not independent test estimates; Phase residual energy is not matched to Clean after reconstruction subtraction.

A. Arm-level CE / accuracy means

Dataset	Architecture	Baseline	Clean	Phase	Extra-Z
CIFAR10-DVS	MaxFormer	0.9470/78.75	0.9504/78.18	0.9530/78.18	0.9510/78.33
CIFAR10-DVS	SEW-ResNet	1.3239/71.92	1.3231/72.04	1.3201/71.95	1.3206/72.34
CIFAR10-DVS	Spikformer	1.0873/68.92	1.1308/67.18	1.0982/68.50	1.0939/68.73
N-Caltech101	MaxFormer	1.9673/73.23	1.9642/72.20	1.9675/72.58	1.9576/72.34
N-Caltech101	SEW-ResNet	3.2329/35.35	3.2373/34.65	3.2305/35.20	3.2316/35.12
N-Caltech101	Spikformer	1.8521/65.51	1.9770/59.85	1.9216/61.10	1.7862/63.11
DVS128Gesture	MaxFormer	0.3503/94.55	0.3947/94.43	0.3759/95.38	0.3598/95.45
DVS128Gesture	SEW-ResNet	0.5173/92.73	0.5246/91.59	0.5199/92.05	0.5242/92.05
DVS128Gesture	Spikformer	0.2628/94.55	0.4315/90.76	0.3833/91.14	0.2828/93.52

B. Paired outcome and control effects

Dataset	Architecture	Base−Clean CE	ΔAcc (pp)	Off−on CE	Phase−Clean CE	Extra-Z−Clean CE
CIFAR10-DVS	MaxFormer	-0.0034 ± 0.0013	-0.5729 ± 0.4060	-0.0031 ± 0.0008	$+0.0026 \pm 0.0005$	$+0.0005 \pm 0.0014$
CIFAR10-DVS	SEW-ResNet	$+0.0009 \pm 0.0043$	$+0.1250 \pm 1.2763$	-0.0031 ± 0.0022	-0.0030 ± 0.0018	-0.0025 ± 0.0032
CIFAR10-DVS	Spikformer	-0.0434 ± 0.0139	-1.7396 ± 1.1878	-0.0350 ± 0.0069	-0.0326 ± 0.0168	-0.0369 ± 0.0170
N-Caltech101	MaxFormer	$+0.0031 \pm 0.0052$	-1.0262 ± 0.5826	$+0.0043 \pm 0.0022$	$+0.0034 \pm 0.0017$	-0.0066 ± 0.0020
N-Caltech101	SEW-ResNet	-0.0044 ± 0.0062	-0.7065 ± 0.3137	$+0.0007 \pm 0.0013$	-0.0068 ± 0.0054	-0.0056 ± 0.0101
N-Caltech101	Spikformer	-0.1249 ± 0.0110	-5.6621 ± 0.7816	-0.0364 ± 0.0191	-0.0554 ± 0.0373	-0.1908 ± 0.0249
DVS128Gesture	MaxFormer	-0.0444 ± 0.0145	-0.1136 ± 1.1607	-0.0401 ± 0.0113	-0.0188 ± 0.0106	-0.0349 ± 0.0072
DVS128Gesture	SEW-ResNet	-0.0073 ± 0.0020	-1.1364 ± 0.4593	-0.0041 ± 0.0032	-0.0047 ± 0.0063	-0.0004 ± 0.0044
DVS128Gesture	Spikformer	-0.1687 ± 0.0652	-3.7879 ± 1.3143	-0.1109 ± 0.0375	-0.0482 ± 0.0077	-0.1487 ± 0.0615

$7/80$ materially mismatched cells, all MaxFormer, have a paired CE or accuracy contrast shift whose within-run sample-bootstrap 95% interval excludes zero. These are descriptive, unadjusted sample intervals and do not measure network-level uncertainty. The original Phase is a dose-mismatched content control; matching its input dose does not match the learned post-adapter injection magnitude.

Capacity and training-budget sensitivity at the CIFAR stem. The fixed b_6 comparison uses SEW and Spikformer, one pilot and one independent network per architecture. Each network has a matched 257-parameter or full-rank adapter crossed with 3 or at most 30 head/adapter epochs. A separate check changes reconstructor training from one to ten epochs under the 257-parameter, 3-epoch setting. The four original arms and sample-ID interventions are retained; each condition uses the same 200 evaluation IDs. Select CE and select reconstruction MSE choose states. The 30-epoch policy permits early stopping, so realized updates differ.



Figure 6: Supplementary first-stem destination-decoder responses for N-Caltech101 and DVS128Gesture. Each matrix averages three networks. Rows show phase-replaced AC source bands and columns show destination bands, with one shared signed CE scale. Positive values denote worse fixed-decoder CE after intervention; negative values denote improvement. Cross-band responses are heterogeneous and are not information-transport estimates.

Table 8: Known-content q CE at complete erasure in all nine cells. One network/adapter instance per cell; Clean accuracy is 100%. The calibration uses the delta-only readout described above, not the natural residual prediction chain.

Dataset	Architecture	Erase	Clean	Phase	Wrong-band	Extra-Z
CIFAR10-DVS	MaxFormer	0.69326	0.00518	0.69345	0.69326	0.69326
CIFAR10-DVS	SEW-ResNet	0.69316	0.00031	0.69325	0.69316	0.69316
CIFAR10-DVS	Spikformer	0.69316	0.00279	0.69316	0.69316	0.69316
N-Caltech101	MaxFormer	0.69322	0.00236	0.69322	0.69322	0.69320
N-Caltech101	SEW-ResNet	0.69315	0.00011	0.69315	0.69315	0.69315
N-Caltech101	Spikformer	0.69315	0.00075	0.69315	0.69315	0.69315
DVS128Gesture	MaxFormer	0.69358	0.02033	0.69333	0.69358	0.69360
DVS128Gesture	SEW-ResNet	0.69299	0.00040	0.69256	0.69299	0.69299
DVS128Gesture	Spikformer	0.69305	0.00286	0.69290	0.69305	0.69305

All 20 within-network sample-bootstrap intervals for the four primary capacity/budget comparisons and their interaction include zero. Longer reconstruction lowers evaluation MSE for both SEW networks and both Spikformer networks, but its Clean CE effect is positive for SEW and negative for Spikformer in both phases. With only two network checkpoints per architecture, this is a narrow sensitivity, not a general architecture-level failure or recovery estimate.

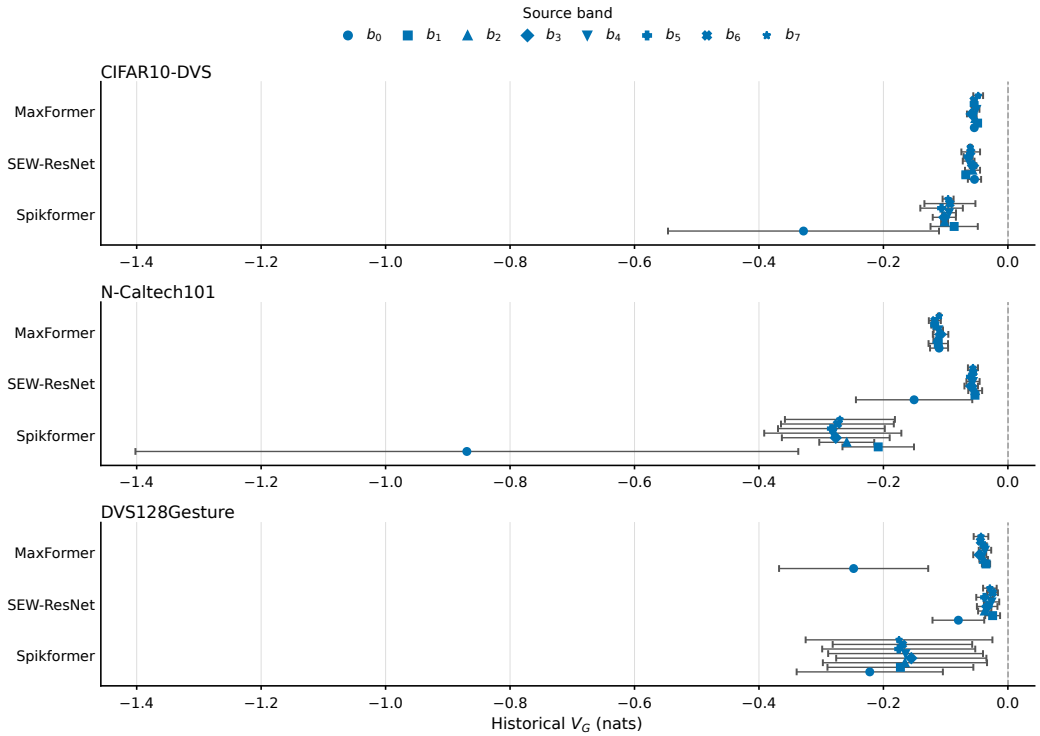


Figure 7: Historical V_G diagnostic in the repaired panel. Each dot is the mean across three network means after averaging two adapters within each network; horizontal intervals show one sample SD across networks. The three dataset blocks share a common horizontal scale and a vertical zero reference; row labels identify architecture and source band. V_G remains negative in all 432 network–adapter–band comparisons, but compares with the per-image better control and is not a loss classifier.

Table 9: CIFAR residual-control CE contrasts in nats, control minus Clean. Mean ± SD across three network means, after averaging two adapter repeats within each network and weighting eight bands equally. Positive values favor Clean. The direct control randomizes the already formed Clean residual; the matched control rescales the original Phase residual.

Architecture	Original Phase	L2-matched Phase	Direct residual phase
MaxFormer	+0.0017 ± 0.0018	+0.0035 ± 0.0009	+0.0025 ± 0.0039
SEW-ResNet	-0.0015 ± 0.0035	-0.0034 ± 0.0042	-0.0016 ± 0.0041
Spikformer	-0.0309 ± 0.0296	-0.0317 ± 0.0300	-0.0330 ± 0.0319

Table 10: CIFAR b_6 sensitivity of Baseline-minus-Clean CE gain (nats). Positive differences favor the first condition named in the contrast. Values are mean ± sample SD over two network-level paired contrasts, one pilot and one independent repeat; they are descriptive, not a network-population interval. Full denotes a dense channel map.

Architecture	Contrast	Difference
SEW-ResNet	Full−257, 3-epoch	+0.0128 ± 0.0054
SEW-ResNet	Full−257, 30-epoch max	-0.0036 ± 0.0103
SEW-ResNet	30−3 max, 257	+0.0095 ± 0.0173
SEW-ResNet	30−3 max, full	-0.0068 ± 0.0016
SEW-ResNet	Recon. 10−1, 257/3	+0.0226 ± 0.0097
Spikformer	Full−257, 3-epoch	+0.0043 ± 0.0170
Spikformer	Full−257, 30-epoch max	-0.0025 ± 0.0180
Spikformer	30−3 max, 257	-0.0299 ± 0.0262
Spikformer	30−3 max, full	-0.0367 ± 0.0272
Spikformer	Recon. 10−1, 257/3	-0.0179 ± 0.0207

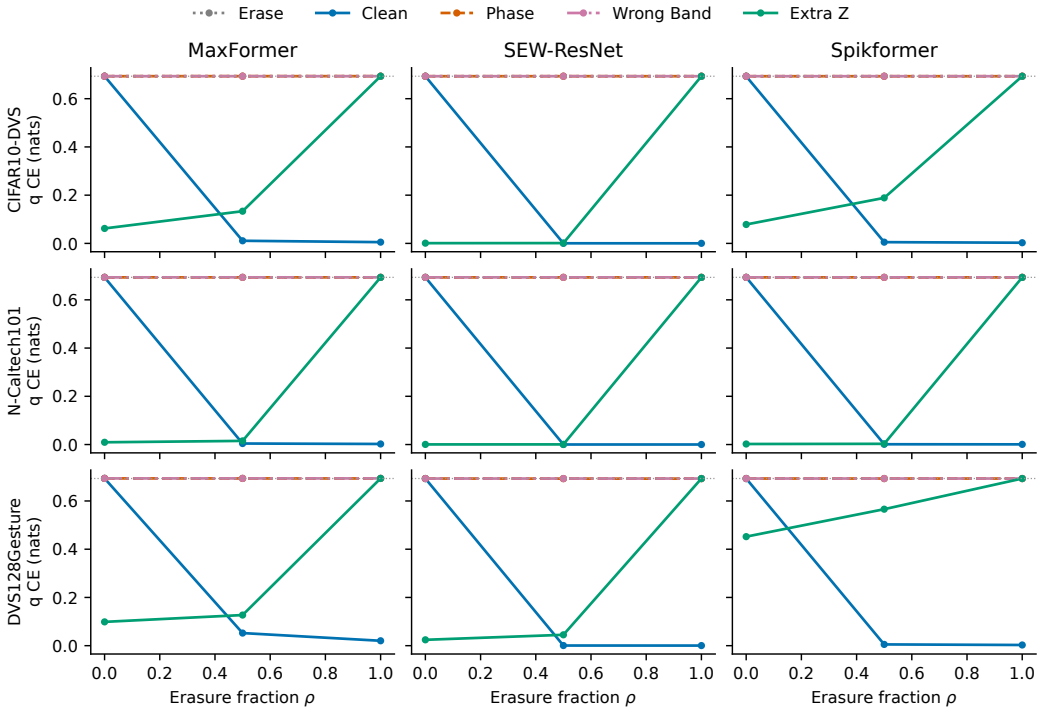


Figure 8: Known-content calibration over the three erasure doses. One fixed network/adapter instance per cell; there is no between-network uncertainty estimate. All arms use the same erased input at each dose. Extra-Z can exploit surviving q at partial erasure. The dotted reference marks uniform binary CE.

G FROZEN-ORIGINAL-HEAD INTERVENTION AND PATCH DETAILS

Three strict train-core MaxFormer checkpoints are evaluated on the same 900 development images, 90 per class. Backbone parameters, original head, and BN statistics are frozen; spiking states reset between forwards. Source phase corruption is fixed by sample ID. The earliest stem LIF output is the patch boundary. No probe, adapter, or reconstructor is fitted. The source and destination AC bands are b_1–b_7; DC sign controls are reported separately. The full-state check includes DC.

In the matched binary domain, $\delta = Z^{\mathrm{clean}} - Z^{\mathrm{corrupt}}$ and $P_c\delta$ ranks eligible corrupt-to-clean bit flips. Exactly the same 10% Hamming budget and post-patch distance from the clean state apply to each sample, source, and requested destination. This preserves the post-LIF binary contract but does not keep the applied increment strictly within band c. The joint target is the union of all non-source AC destinations and is evaluated by a separate downstream forward, not a sum of single-band CE effects. Phase controls use the same legal flip budget; zero replacement reproduces corrupt predictions. Complete replacement of the captured post-LIF state reproduces clean logits, CE, and predictions exactly in all $3 \times 900 \times 8$ audited forwards.

Table 11: Original-head AC patch recovery (nats), corrupted minus patched CE. Each value is mean $\pm$ SD across three frozen-network summaries. The binary domain fixes the 10% flip budget; Fourier hybrid patches inject continuous states and unequal energy. Joint is an actual forward over all non-source AC destinations. Domains are not pooled.

Domain	Same AC	Single cross AC	Joint non-source AC
Binary	+0.0221 $\pm$ 0.0072	+0.0115 $\pm$ 0.0035	-0.0036 $\pm$ 0.0032
Hybrid	+0.0395 $\pm$ 0.0069	+0.0044 $\pm$ 0.0029	+0.0243 $\pm$ 0.0064

The original Fourier hybrid has a positive joint recovery on all three networks, but its mean joint increment energy (0.05657) is about six times the single cross-band increment (0.00943). Its same/cross patches also have nonbinary post-LIF fractions of $0.367/0.354$. Hence this larger joint response cannot be attributed to multi-band complementarity. The equal-budget binary joint result is negative on all three networks; it does not exclude other budgets or interventions. Phase controls can themselves harm the classifier, and the binary oracle-versus-phase advantage is small and heterogeneous. A patch effect under source corruption does not estimate conserved or missing task information.

H HISTORICAL LOCALIZATION, CALIBRATION, AND RECOVERY ANALYSES

These experiments explain the mechanism and sensitivity of the diagnostic under their recorded conditions. Their checkpoints, splits, boundaries, and statistics differ from the unified panel and are not pooled with it.

H.1 HISTORICAL OPERATOR LOCALIZATION AND DATASET COMPARISONS

The main-text trace (Figure 2) and the complete matrix below come from the same historical MaxFormer artifact, not independent runs. The latter retains all AC source bands and boundary responses.

At embedding output, the input-b_5 intervention still changes the b_0 fixed-decoder CE more than the b_5 decoder CE (0.374 versus 0.267 nats). These responses may include distribution shift and do not reconstruct source content.

The frequency locations are task dependent. In N-Caltech101, input b_2/b_3 have same-band shares of $23.6/22.8\%$ at the first LIF and $12.8/11.6\%$ at embedding output, where low bands receive $45.1/38.1\%$ of their influence. DVS128Gesture shows the same qualitative early transition despite high baseline accuracy. SEW-ResNet exhibits same-frequency propagation through stem Conv+BN and dispersion after the first PLIF (Figure 10). Its most salient band is not the band with the greatest relative off-diagonal influence: b_5 is highly sensitive, whereas b_1 routes more strongly across frequencies at the first block. Saliency, energy share, and routing are distinct axes.

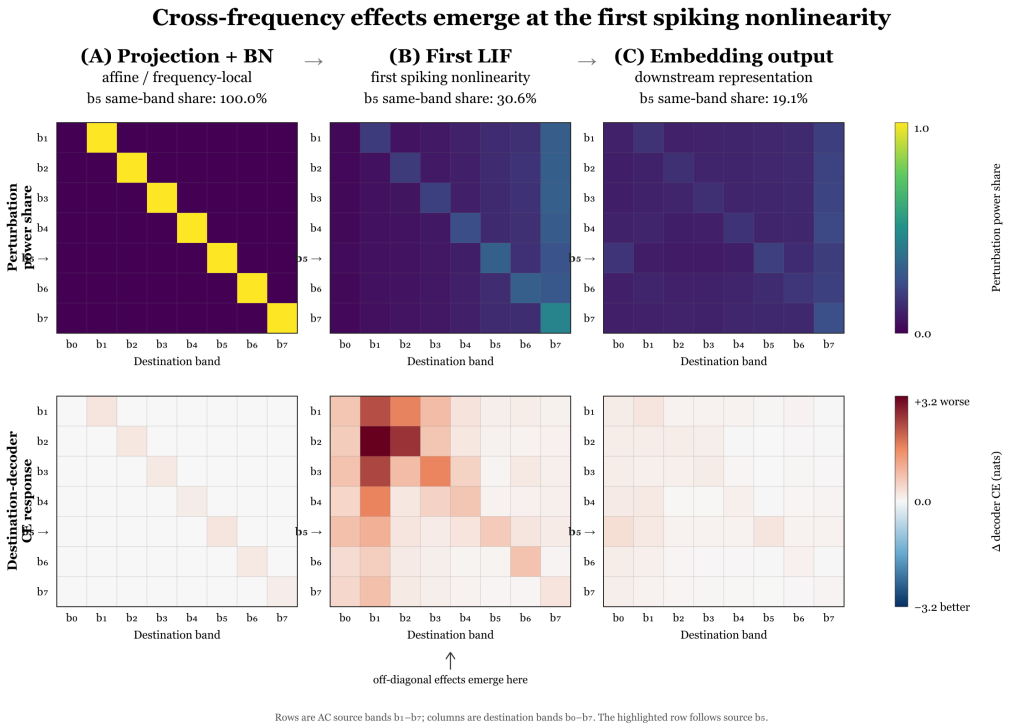


Figure 9: Complete historical MaxFormer operator trace. The three columns follow the same input-band intervention from projection and evaluation-mode BN through the first LIF to the embedding output. Top: destination shares of perturbation power on a shared $[0, 1]$ scale; the marked b_5 source row has same-band shares of 100.0%, 30.6%, and 19.1%. Bottom: destination-decoder CE responses on a shared symmetric ± 3.22-nat scale. Rows are AC source bands b_1–b_7 and columns are destination bands b_0–b_7; DC is excluded because ordinary phase shuffling does not identify its source intervention. Off-diagonal responses alone do not establish information transport or preservation. Historical 9000/1000 analysis, one backbone checkpoint, three phase and three projection seeds.

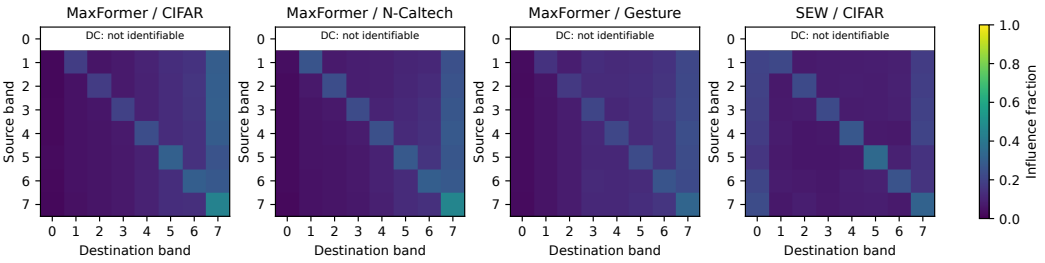


Figure 10: First-spike influence matrices across three MaxFormer datasets and the SEW CIFAR10-DVS case, plotted on the same scale. The common observation is cross-frequency redistribution, not identical routing weights or a universal destination band. DC source rows are excluded. Samples and checkpoints differ between panels; these are descriptive architecture/task comparisons, not pooled independent replicates.

H.2 Attention-level case study

In Spikformer-2-256, we trace 13 boundaries around the first spiking self-attention block using a fixed checkpoint and an 800/200 train-side split. Q/K/V thresholding slightly increases the aggregate same-band share in this already mixed representation. The subsequent $QK^\top V$ interaction increases

off-diagonal share by approximately 0.40–0.46 percentage points relative to the individual spiking branches. The post-attention LIF aggregate change is close to zero. This does not support a rule that every spike layer monotonically increases cross-frequency mixing.

The final block contains broad perturbation effects and destination decoders sensitive to off-target source bands. However, raw classifier CE increases do not pass the historical eight-band Holm criterion in this small screening set. We therefore treat this case as operator redistribution and fixed-readout evidence, rather than equally strong task-level confirmation for all source bands. Moreover, these upstream interventions affect multiple attention branches: differences between boundary matrices do not isolate an independent causal contribution from Q, K, or V.

H.3 Graded calibration and alternative explanations

The historical MaxFormer calibration varies carrier amplitude $\lambda \in \{0.05, 0.1, 0.2\}$ and erasure fraction $\rho \in \{0, 0.25, 0.5, 0.75, 1\}$ with three adapter seeds. Ten of twelve nonzero-erasure points pass the recorded operational gate; the three zero-erasure points are controls. The smallest passing tested point is $(0.05, 0.25)$, not a universal detection limit. A state-distance audit at two points finds functional q rescue without movement toward the No-Erase representation. It does not establish faithful reconstruction or extend to all grid points. These historical results are distinct from the three-architecture q calibration in Appendix F.1.

More spectral delivery need not improve prediction. In a historical CIFAR10-DVS late bypass, an oracle original-band path with a concatenation head increases the head's b_5 energy share from 2.91% to 8.25% in a same-weight intervention, a relative increase of 183.5%. Test accuracy is 82.50%, versus 82.80% for the reference model. Turning the learned bypass on rather than off changes accuracy by only 0.20 percentage points and slightly worsens CE. Because the band was chosen using historical test analysis, this experiment is an exploratory manipulation check, not a leakage-free method comparison. It demonstrates that energy delivery itself is an insufficient success criterion, without identifying the unique reason the classifier failed to benefit.

A narrow downstream projection can manufacture a residual. Conditional ridge/CountSketch probes find no robust correct-residual advantage beyond Extra-Z at the preregistered MaxFormer boundary. Apparent gains with downstream dimension 256 disappear when it is increased to 512. The SEW conditional probe similarly loses its small positive effect under capacity controls. The supported conclusion is restricted to these representations, projections, and readouts: an apparent sufficiency gap may reflect omitted downstream features. It is not evidence that the full network is lossless.

H.4 Historical natural rescue and recovery variants

The historical natural-rescue screens use different training and evaluation conditions from the unified panel. MaxFormer full-sample validation covers eight bands, three network seeds, and two adapter seeds: none of the 48 band–adapter combinations passes the historical provisional gate. The historical SEW full-sample fresh-head assay fails readout health and cannot support a natural No-Go conclusion. Spikformer produces historical provisional development positives on CIFAR10-DVS b_0 and N-Caltech101 b_0/b_1, but their dependence on the protocol prevents attributing the gains specifically to recovery of lost information.

The CIFAR b_0 fresh-head follow-up used the same historical SSA boundary, candidate band, $8100/900$ development split, and two network by two adapter seeds at both budgets. The three-epoch mean CE and V_G gains reverse at 30 epochs (Table 12); all four long runs have negative same-weight on/off CE gain. Their small positive accuracy gains show that CE and classification need not move together. Selection uses the same validation data, so this is a within-protocol sensitivity comparison. The unified 3×3 experiment instead measures the earlier stem spike and is not validated by the long SSA budget.

Retaining the pretrained head also fails to establish recovery: the corrected adapter-only runs select the zero-adapter source state. Matched fine-tuning produces additional baseline drift and side-path harm, making its negative outcomes unsuitable as universal architecture No-Go evidence. Source and function-preserving constraints reduce some training-health concerns but do not yield a stable

Table 12: Historical short/long fresh-head comparison on Spikformer/CIFAR10-DVS b_0 at the SSA interaction-to-spike boundary. Each entry is mean $\pm$ descriptive SE of two network means, each averaging two adapter seeds. CE gain is Baseline-minus-Clean; accuracy is Clean-minus-Baseline in percentage points. The same candidate and development split are reused. Budget changes selected optimization trajectories; these are not independent test estimates.

Head/adapter budget	CE gain	ΔAcc	$\overline{V}_G$
3 epochs	$+0.0689 \pm 0.0103$	$+5.75 \pm 1.03$	$+0.0601 \pm 0.0052$
30 epochs	-0.0727 ± 0.0205	$+0.36 \pm 0.03$	-0.1028 ± 0.0117

cross-split gain. We report these as limits of the tested recovery systems, not proof of zero conditional information.

I EVIDENCE PROVENANCE

The accompanying manuscript assets are regenerated from saved JSON results. The historical four-operator records remain in the source archive; Figure 2 uses the CIFAR MaxFormer operator-lineage artifact; Figure 10 adds the corresponding N-Caltech, Gesture, and SEW all-band artifacts. Table 12 aggregates the source network and adapter records from the existing audited manifest. Two entries in the historical run ledger had swapped numeric summaries; the directory names and saved configurations themselves agree, and the manuscript table follows the actual artifacts. The unified assets use a separate manifest, generated by `scripts/build_unified_paper_assets.py`. It records hashes of the downloaded server summaries and per-image arrays, verifies canonical utility for all 144 instance–band pairs, and reads only v2 lineage results. The supplementary review manifest verifies 96 operator records and recomputes all three 900-image patch panels from saved logits, including CE, predictions, zero equivalence and patch/control energy matching. The source manifests record full SHA-256 hashes and distinguish missing canonical utility fields from observed negative values.

This preparation audit verifies saved records, split counts, source checkpoint hashes for the R1 family, and the arithmetic used to construct figures and tables. It does not regenerate every activation, retrain every network, recompute all historical bootstraps, or establish an exhaustive novelty claim. The operator identities and counterexamples have a separate CPU numerical check. Scientific interpretation remains bounded by the recorded protocols and exposure history.