SPECTRAL ACCESSIBILITY: PROBING WHAT EEG FOUNDATION MODELS PRESERVE

Anonymous authors
Paper under double-blind review

ABSTRACT

Downstream accuracy measures task utility, but does not reveal how broadly a frozen EEG representation exposes signal structure. We introduce *spectral accessibility*, a label-independent audit of a declared second-order state built from multiresolution complex cross-spectral matrices. A subject-disjoint fit selects and whitens 64 reliable state directions. Their recoverability defines an accessibility spectrum; SPA-20 tests broad coverage by averaging its weakest 20% under the worst declared input environment. We distinguish the unrestricted population operator, its best-linear counterpart, and the finite cross-fitted ridge estimate. Population theory gives data-processing inequalities and task-error bounds, including bounds restricted to accessible subspaces. Across five masked-reconstruction or masked-prediction EEG foundation models, seven datasets, and four input conditions, all 140 estimated linear lower tails reach zero. Strong split-taper controls on four datasets show that the same pipeline can recover the target's weak directions, although pooling, subject transfer, and probe estimation remain possible bottlenecks. Non-zero access is concentrated in upper subspaces. We report their exploratory environmental profiles and conditional uncertainty. In this regime, SPA-20 is a diagnostic stress test; the full spectrum, null floor, and split-taper reference carry the comparative information.

1 INTRODUCTION

An EEG foundation model maps a multichannel recording X to a representation $Z = F_\theta(X)$. Most evaluations ask whether this representation can support a downstream label: sleep stage, motor imagery, seizure, or diagnosis. The answer depends on the checkpoint, dataset, decoder, adaptation rule, and optimization budget. Frozen probes and fine-tuning can consequently give different model orderings (Kornblith et al., 2019; He et al., 2022; Xiong et al., 2025). Task performance alone cannot distinguish broad signal preservation from access to a small task-relevant subspace.

We ask a complementary question: given a declared signal state $S = \Phi(X)$, which directions in that state can be recovered from a frozen representation? The target is computed from the recording, without downstream labels or encoder updates. The choice of state makes the question scientifically specific. There is no task-free ordering of representations independent of which signal properties matter.

Our state describes second-order EEG structure. Complex cross-spectral density (CSD) matrices retain channel power and cross-channel amplitude and phase relationships across frequencies and time windows. Concatenation yields 2,550–84,125 raw coordinates per epoch. We first project these coordinates onto at most 64 PCA modes that pass a split-taper reliability screen, then whiten them using a separate subject cohort. All subsequent covariance operators act in this retained space, not in the raw feature space.

The accessibility operator measures explained standardized variance jointly across those directions. Its eigenvalues reveal whether access is spread broadly or concentrated in a few directions. SPA-20 is a stress test for broad access: it takes the weakest-20% mean under the worst declared environment. A zero estimate identifies a boundary of this assay, not complete information loss or a model ranking. The primary diagnostic outputs are therefore the spectrum, its null floor, and a split-taper reference.

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.

We make three contributions. First, we define unrestricted and linear population accessibility, state their distinct invariances, and relate spectral coverage to error on tasks mediated by the state. The lower-tail risk bound gives no improvement over the zero-state baseline at zero coverage; a subspace bound retains meaning for tasks aligned with accessible directions. Second, we implement a subject-disjoint, nested ridge assay with reliability screening, full residual covariance, derangement controls, and split-taper references. Third, we audit five predominantly masked-autoencoding or masked-prediction models on seven datasets and four input views. All 140 lower-tail estimates are zero, while upper subspaces remain accessible. An explicitly exploratory analysis describes these subspaces across environments and reports their uncertainty. Code and result artifacts are available in the anonymous implementation repository.[1]

2 Related work

EEG foundation models and evaluation. Our empirical scope is masked reconstruction and prediction. REVE, CSBrain, and CBraMod use masked reconstruction; EEGPT combines masked reconstruction with representation alignment; LaBraM predicts discrete codes learned through Fourier-spectrum reconstruction (El Ouahidi et al., 2025; Zhou et al., 2025; Wang et al., 2025; 2024; Jiang et al., 2024). These objectives differ, so the cohort is not a controlled comparison of reconstruction losses. Contrastive BENDR (Kostas et al., 2021) and the cross-data tokenization framework BIOT (Yang et al., 2023) are relevant alternatives, but have no complete SPA-20 results in this audit. Their spectral profiles could differ; our findings do not extend to all self-supervised EEG models. Existing benchmarks examine frozen probes, adaptation, and multiple evaluation views (Xiong et al., 2025; Wu et al., 2025; Yang et al., 2026; Liu et al., 2026; Kommineni et al., 2026b). We add a signal-state audit to those task-based evaluations.

Probing and transfer. Linear evaluation is widely used to separate frozen representation content from end-to-end adaptation. Its interpretation depends on pretraining: supervised vision models can show strong probe–transfer agreement, whereas masked autoencoders may expose useful features only after nonlinear or partial adaptation (Kornblith et al., 2019; He et al., 2022). We therefore treat the decoder family as part of the estimand and call the implemented quantity *linear* spectral accessibility. The population theory uses the unrestricted conditional mean; the experiment uses ridge regression and does not equate failure of a linear readout with absence of information.

Spectral structure. Multitaper estimation reduces leakage and yields multiple views of the same spectrum (Thomson, 1982). CSD matrices retain both per-channel power and cross-channel phase structure; a log-Euclidean map supplies stable Euclidean coordinates for positive-definite matrices (Arsigny et al., 2006). Recent work identifies low-frequency and aperiodic biases in reconstruction-based EEG foundation models (Kommineni et al., 2026a; Yu et al., 2026). Our contribution is complementary: a multivariate spectrum of how much time-, frequency-, channel-, and cross-channel state is accessible through a frozen checkpoint.

3 Spectral accessibility

3.1 A declared state and checkpoint view

Let X be a clean EEG epoch under a fixed sampling law and let $S = \Phi(X) \in \mathbb{R}^D$ be a declared state. A state-fit sample, disjoint by subject from all scoring data, determines a reliable q-dimensional subspace and a whitening transform. The resulting target $\bar{S} \in \mathbb{R}^q$ satisfies $\mathbb{E}[\bar{S}] = 0$ and $\mathrm{Cov}(\bar{S}) = I_q$ in the population idealization. Whitening prevents high-variance delta or scale directions from dominating the score.

Let $\mathcal{E}$ be a finite, declared set of input transformations applied to the same epoch distribution. Environment e changes the checkpoint view,

$$Z_{\theta,e} = F_\theta(T_e(G_\theta(X))). \tag{1}$$

Here G_θ is the input adapter and F_θ includes the encoder and pooling operation. The state map Φ includes the fixed canonical target route. The target always comes from clean X, and its law is the

[1] https://anonymous.4open.science/r/UniformEvalBench-78C6/

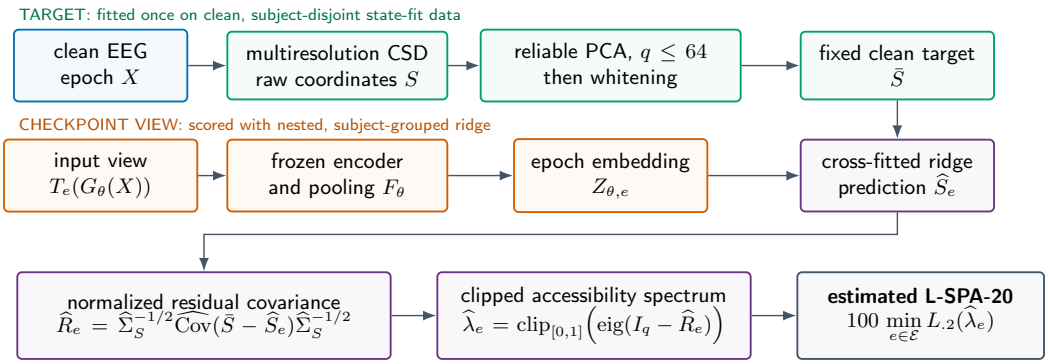

Figure 1: The finite spectral-accessibility assay. Clean state-fit subjects determine PCA, reliability screening ($q \leq 64$), and whitening. Frozen checkpoint views predict the same clean target on score subjects. Scoring rewhitens residual covariance by the empirical target covariance, computes the spectrum, clips negative eigenvalues, and averages the weakest 20% before taking the environmental minimum.

same for every e. We condition on the fitted state map; exact identity covariance is a population idealization, not a claim about finite Ledoit–Wolf whitening.

3.2 The accessibility operator

Under squared loss, the population-optimal prediction is the conditional mean

$$m_{\theta,e}(Z_{\theta,e}) = \mathbb{E}[\bar{S} \mid Z_{\theta,e}, e]. \tag{2}$$

We define the accessibility operator

$$A_{\theta,e} = \mathrm{Cov}(m_{\theta,e}(Z_{\theta,e}) \mid e). \tag{3}$$

For any unit state direction v,

$$v^\top A_{\theta,e} v = \mathrm{Var}\big(\mathbb{E}[v^\top \bar{S} \mid Z_{\theta,e}, e] \mid e\big) = 1 - \mathbb{E}\big[\mathrm{Var}(v^\top \bar{S} \mid Z_{\theta,e}, e) \mid e\big]. \tag{4}$$

Thus $0 \preceq A_{\theta,e} \preceq I_q$. Its ordered eigenvalues $0 \leq \lambda_{e,(1)} \leq \cdots \leq \lambda_{e,(q)} \leq 1$ form the *accessibility spectrum*. Eigenvectors identify state directions; eigenvalues measure their recoverability. This is a second-order conditional-mean criterion: $A = 0$ does not imply statistical independence.

For the experiment, define $m^{\mathrm{lin}}_{\theta,e}$ as the best affine least-squares predictor from $Z_{\theta,e}$, and

$$A^{\mathrm{lin}}_{\theta,e} = \mathrm{Cov}(m^{\mathrm{lin}}_{\theta,e}(Z_{\theta,e}) \mid e), \qquad 0 \preceq A^{\mathrm{lin}}_{\theta,e} \preceq A_{\theta,e} \preceq I_q. \tag{5}$$

These are population objects. Cross-fitted ridge estimates the linear object only under suitable consistency conditions; its finite residual operator need not be positive semidefinite.

3.3 SPA-20: a lower-tail functional

For tail mass $\alpha \in (0, 1]$, set $a = \alpha q$, $k = \lfloor a \rfloor$, and $\delta = a - k$. Define the fractional lower-tail mean

$$L_\alpha(\lambda) = \frac{\sum_{j=1}^{k} \lambda_{(j)} + \delta \lambda_{(k+1)}}{\alpha q}. \tag{6}$$

The fractional term is omitted when $\delta = 0$, including $\alpha = 1$; an empty sum is zero. The general score and operating point are

$$\mathrm{SPA}_\alpha(\theta) = 100 \min_{e \in \mathcal{E}} L_\alpha(\lambda(A_{\theta,e})), \qquad \text{SPA-20}(\theta) \equiv \mathrm{SPA}_{0.2}(\theta). \tag{7}$$

Replacing A by A^{lin} defines population L-SPA-α. We use $\alpha = .2$ as the primary operating point. At $q = 64$, the tail contains 12 full eigenvalues and weight .8 on the 13th. It detects weak subspaces, but a positive tail average is not a positive lower bound on every individual direction.

3.4 POPULATION PROPERTIES

Bounds and endpoints. Equation 4 gives $\mathrm{SPA}_\alpha \in [0, 100]$. Independence of $Z_{\theta,e}$ and $\bar{S}$ yields $A_{\theta,e} = 0$. Perfect state recovery, $\bar{S} = m_{\theta,e}(Z_{\theta,e})$ almost surely, is equivalent to $A_{\theta,e} = I_q$. These statements concern recovery of the declared state, not sufficiency for arbitrary labels.

Representation data processing. If $Z'_e = h(Z_{\theta,e})$ for measurable deterministic h, then

$$A_{Z',e} \preceq A_{\theta,e}, \qquad \lambda_j(A_{Z',e}) \le \lambda_j(A_{\theta,e}) \quad \forall j. \tag{8}$$

This unrestricted population theorem covers pooling, projection, and quantization; an almost-surely invertible measurable map with measurable inverse preserves A. For A^{lin}, the analogous claims apply to affine maps and invertible affine coordinate changes. Arbitrary nonlinear maps can improve linear access. Finite ridge probes need not obey either monotonicity: a lossy projection may reduce estimation variance. Isotropic ridge is exactly invariant to orthogonal embedding rotations at any penalty with otherwise equivariant preprocessing and fixed folds. Our coordinatewise standardization breaks that equivariance. Separate input perturbations are not ordered by this theorem because one encoded view need not be a function of another.

Target-coordinate invariance. Any invertible linear change of retained target coordinates, followed by rewhitening, rotates $\bar{S}$ by an orthogonal matrix Q. Therefore $A'_{\theta,e} = QA_{\theta,e}Q^\top$, leaving the spectrum unchanged. The same holds for A^{lin}. This is exact population rewhitening within a fixed subspace; it does not make PCA selection, shrinkage whitening, or the whole finite pipeline invariant to arbitrary raw-state reparameterization.

Implication for mediated tasks. Let $Y = g_e(\bar{S}) + \varepsilon$, with g_e L_e-Lipschitz and $\mathbb{E}[\varepsilon \mid \bar{S}, Z_{\theta,e}, e] = 0$. For $s = \mathrm{SPA}_\alpha(\theta)/100$,

$$\mathbb{E}\|\bar{S} - m_{\theta,e}(Z_{\theta,e})\|_2^2 = q - \mathrm{tr}(A_{\theta,e}) \le q(1-s), \tag{9}$$

$$\mathbb{E}|g_e(\bar{S}) - g_e(m_{\theta,e}(Z_{\theta,e}))|^2 \le L_e^2\{q - \mathrm{tr}(A_{\theta,e})\} \le L_e^2 q(1-s). \tag{10}$$

At $s = 0$, the last bound equals the $L_e^2 q$ bound obtained by predicting the zero state. Positive s improves this generic bound, but may still be loose for a particular scalar task. For $g_e(s) = h_e(Ps)$, where $PP^\top = I_r$ and h_e is L_e-Lipschitz, the sharper bound is

$$\mathbb{E}|h_e(P\bar{S}) - h_e(Pm_{\theta,e})|^2 \le L_e^2 \,\mathrm{tr}\big(P(I_q - A_{\theta,e})P^\top\big). \tag{11}$$

It can be small on an accessible subspace despite a zero lower tail. The same statements hold with m^{lin}, A^{lin}, and linear SPA. They are population bounds, not empirical guarantees obtained by substituting clipped estimates. Appendix A proves these statements and clarifies the noise baseline.

4 A MULTIRESOLUTION SPECTRAL STATE

For each epoch with C channels, we estimate CSD matrices over five bands: δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–45 Hz). A fine resolution uses 2-s windows with 1-s steps; a coarse resolution covers the full epoch. Six discrete prolate spheroidal tapers with time-bandwidth product 3.5 are averaged. Band averaging increases the available rank but does not ensure full rank: with K tapers and B_b in-band bins, $\mathrm{rank}(C_b) \le \min(C, KB_b)$. Narrow bands at high channel counts therefore require diagonal loading (Appendix C.3).

For each regularized Hermitian CSD $C_{b,r,t}$, we separate total scale from normalized shape:

$$s_{b,r,t} = \left[\log \mathrm{tr}(C_{b,r,t}),\ \mathrm{svec}\left(\log \frac{C_{b,r,t}}{\mathrm{tr}(C_{b,r,t})}\right)\right]. \tag{12}$$

Here svec stores the real diagonal and $\sqrt{2}$-scaled real and imaginary upper-triangular terms. It is a norm-preserving vectorization of a Hermitian matrix into C^2 real coordinates. Unit-trace shape has only $C^2 - 1$ degrees of freedom, so adding scale gives an overcomplete $C^2 + 1$-coordinate representation. If $H = \log(C/\,\mathrm{tr}\,C)$, the constraint is $\mathrm{tr}(\exp H) = 1$, not $\mathrm{tr}\,H = 0$, before numerical eigenvalue flooring. Concatenating bands, times, and resolutions yields

$$D = (C^2 + 1) \times 5 \times (N_{\mathrm{fine}} + 1), \tag{13}$$

which ranges from 2,550 for four-channel HMC to 84,125 for 58-channel MIMUL.

State fitting uses only the subject-disjoint fit partition. We compute the state three times: all six tapers and two disjoint halves, tapers $\{0, 2, 4\}$ and $\{1, 3, 5\}$. PCA is fitted to the mean of the halves with a prespecified cap of 64 modes. A mode is retained only if its split-taper correlation is at least .70. In all seven datasets, all 64 candidate modes pass; minimum correlations range from .970 to .999. The full-taper coordinates are then centered and whitened using a Ledoit–Wolf covariance estimate fitted on state-fit subjects. The split halves are transformed by the same fixed map. Disjoint taper sets are two views of one epoch, not independent biological measurements; high fit-partition correlation does not by itself establish cross-subject recoverability.

5 Finite estimator and experimental design

5.1 Cross-fitted linear accessibility

State fitting and checkpoint scoring use disjoint subjects. For each input view, nested subject-grouped cross-validation predicts the fixed clean target $\bar{S}$ from $Z_{\theta,e}$. The outer loop has up to five folds and the inner loop up to four. Each training fold fits a coordinatewise StandardScaler and an affine multi-output ridge model. A *single scalar penalty*, shared across all 64 outputs, is chosen from 20 log-spaced values in $[10^{-3}, 10^{4}]$ by mean squared error averaged over target coordinates. Only outer-held-out predictions are concatenated.

Let $\widehat{\Sigma}_S = \widehat{\mathrm{Cov}}(\bar{S})$ on the scoring epochs. This covariance is common to every input environment. Define

$$\begin{aligned} \widehat{\Sigma}_{\mathrm{res},e} &= \widehat{\mathrm{Cov}}(\bar{S} - \widehat{S}_e), \\ \widehat{B}_e &= I_q - \widehat{\Sigma}_S^{-1/2} \widehat{\Sigma}_{\mathrm{res},e} \widehat{\Sigma}_S^{-1/2}, \qquad \widehat{\lambda}_e = \mathrm{clip}_{[0,1]}(\lambda(\widehat{B}_e)). \end{aligned} \tag{14}$$

Because residual covariance is positive semidefinite, $\widehat{B}_e \preceq I_q$: only the lower clip can be active in exact arithmetic. Negative eigenvalues mark directions whose centered prediction-error variance exceeds target variance. Clipping maps them to zero; it does not establish population inaccessibility. We call $100 \min_e L_{.2}(\widehat{\lambda}_e)$ the estimated L-SPA-20.

Joint penalty selection can favor well-predicted directions and leave weak ones poorly regularized. No coordinate-specific penalty experiment was saved, so its effect is unresolved. Such tuning would also depend on the chosen target basis. The covariance score removes residual means and therefore is not itself mean squared prediction error; Appendix A makes this distinction explicit.

5.2 Controls

An identity oracle predicts the target from itself. Subject-block derangements pair each recipient with a different donor subject to estimate an empirical null floor; no subject maps to itself. Each model–dataset–environment cell uses 20 derangements. The split-taper control predicts one half-taper state from the other with the same grouped ridge pipeline; its null uses five derangements. This control measures both measurement agreement and cross-subject estimator generalization, so it is not a biological ceiling.

Rotation and PCA32 are diagnostics of the implemented probe. Coordinatewise standardization can change after rotation, and finite ridge can improve after dimension reduction. Moreover, the saved PCA32 control fits PCA on all scoring embeddings before grouped cross-validation. It is therefore a transductive, target-free diagnostic, not a fully nested generalization estimate. Clean-input uncertainty uses a leave-one-subject-out jackknife that reruns the probe while holding the fitted target map and baseline quantities fixed.

5.3 Models, datasets, and environments

We evaluate five masked-reconstruction or masked-prediction checkpoints with complete four-condition results: EEGPT (Wang et al., 2024), REVE (El Ouahidi et al., 2025), CSBrain (Zhou et al., 2025), CBraMod (Wang et al., 2025), and LaBraM (Jiang et al., 2024). The seven public datasets

Table 1: State construction and split-taper references. Each dataset contains 20 model–environment cells. The lower-tail reference measures cross-subject recovery of one taper view from the other through the same estimator.

Dataset	fit/score subj.	fit/score epochs	raw D	q	rel. min	ref. $L_{\cdot 2}$
HMC	30/30	1200/1200	2,550	64	0.972	0.852
Motor-MV	30/30	1200/1200	72,020	64	0.998	0.843
MIMUL	12/13	480/520	84,125	64	0.999	0.780
ADFTD	30/30	1190/1200	18,100	64	0.982	0.736
Workload	18/18	720/715	7,240	64	0.970	0.424
Siena	7/7	280/280	36,500	64	0.998	0.354
BCIC-IV-2a	4/5	160/200	9,700	64	0.991	0.019

cover sleep, dementia, motor imagery/execution, seizure detection, workload, and upper-limb movement. Table 1 reports the exact state and split sizes; Appendix C gives dataset citations, preprocessing, adapters, and checkpoint details.

The four checkpoint views are clean input; per-epoch zeroing of $\lfloor C/4 \rfloor$ channels; decimation to 100 Hz followed by linear interpolation to 200 Hz; and colored channel noise at 10 dB signal-to-noise ratio. The implementation scales Fourier amplitude as $1/f$, hence power as $1/f^2$. The clean target and every target-side transformation remain fixed. We cap both partitions at 30 subjects and sample at most 40 evenly spaced epochs per subject. ADFTD has 10 fewer fit epochs and Workload five fewer score epochs than the cap because some subject records supply fewer eligible epochs. Covariances remain epoch-weighted.

Appendix B gives the algorithm; Appendix C specifies preprocessing, adapter routing, pooling, and duration handling. The audit evaluates these complete checkpoint views, including their input and pooling choices.

6 Results

6.1 The lower tail reaches the estimator boundary

Across all $5 \times 7 \times 4 = 140$ model–dataset–environment cells, L-SPA-20 is exactly zero. Under clean input, the clipped median eigenvalue is also zero in 34 of 35 cells; REVE/HMC is the only exception, at .0086. Yet the upper spectrum is non-zero (Figure 2). The estimated linear spectrum is therefore concentrated in a smaller subspace of the retained state. This conclusion concerns the pooled checkpoint view and finite probe.

Fit-partition split-taper correlations are high on every dataset (minimum .970–.999), but this is different from predicting a taper view in new subjects. On HMC, Motor-MV, MIMUL, and ADFTD, the latter control recovers the weak tail at .736–.852. Thus the finite pipeline can recover these target directions from another spectral view; target noise alone is an inadequate account of the checkpoint-view zeros. Workload, Siena, and BCIC-IV-2a have lower references (.424, .354, .019). Their controls also face subject shift and estimation variance, especially with only seven or five score subjects. The artifacts do not isolate these causes, and low reference scores do not establish physical unrepeatability.

All four environments tie at zero. This does not establish robustness: the lower-tail statistic has reached its boundary. The exploratory upper-tail profiles below and in Appendix E resolve environmental differences using the existing results.

6.2 Controls separate population theory from estimator behavior

The oracle reaches L-SPA-20 $\approx$ 100, while derangements reach zero in the *lower tail*. For EEGPT/HMC, rotation changes the clipped spectral mean from .20838 to .20594. This is compatible with the pipeline's non-equivariant standardization, not an exception to ridge's orthogonal invariance under equivariant preprocessing. PCA32 increases both mean and upper-tail scores on ADFTD, Siena, and BCIC-IV-2a (Appendix D). Variance reduction can outweigh information loss in

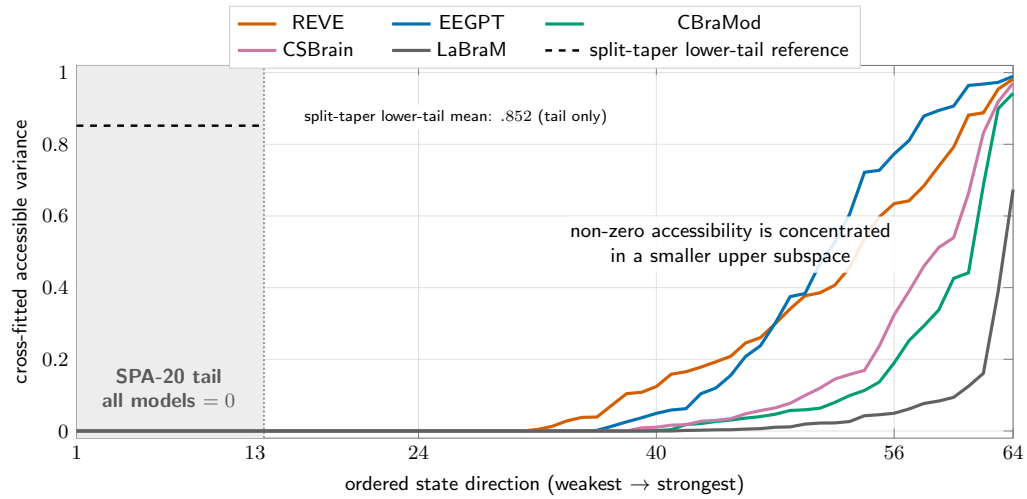

Figure 2: Clean-input accessibility spectra on HMC. Eigenvalues are ordered from the weakest to the strongest recoverable direction. The shaded region is the prespecified SPA-20 tail: all five model estimates are at zero there. The models nevertheless expose non-zero, and in some cases high-accessibility, directions in a much smaller upper subspace. The dashed segment shows the .852 split-taper lower-tail mean only within the shaded tail. The split-taper upper-tail mean is .996; the segment is not a ceiling across all directions.

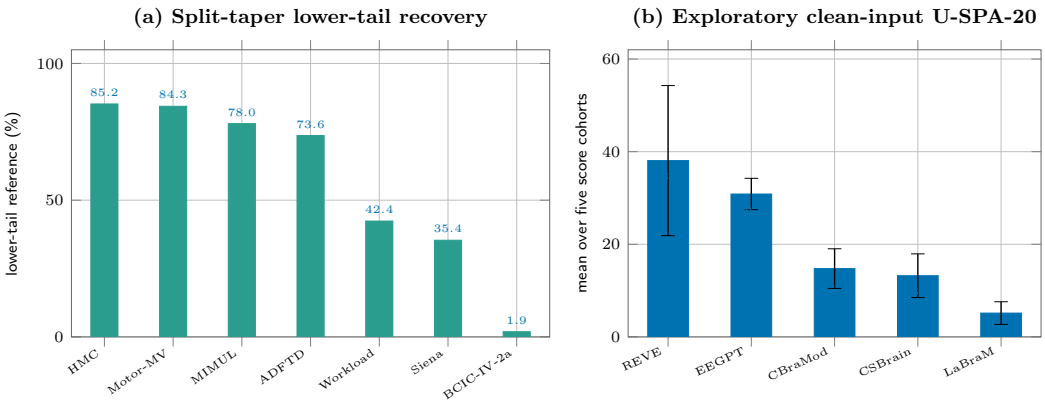

Figure 3: Empirical summary. **(a)** Split-taper lower-tail recovery through the grouped ridge pipeline, alongside 20 zero model–environment lower tails per dataset. This reference reflects measurement agreement and cross-subject estimation. **(b)** Exploratory clean U-SPA-20 means over five cohorts with at least 13 score subjects. Error bars are approximate conditional 95% jackknife intervals, propagated assuming independent datasets; state-map, null-floor, and repeatability uncertainty are omitted.

a finite probe; the transductive PCA fit is an additional caveat. Identical zero lower tails cannot test monotonicity.

Finite spectra also expose why raw scalar comparisons are unsafe. Model-specific null upper tails vary substantially with embedding width, sample size, and target dimension; on Motor-MV they range from approximately .09 to .30. The plain spectral mean is rank-unstable to clipping: its clipped-versus-raw Kendall correlation is negative on six of seven datasets. By contrast, the upper-tail family has correlation $+1$ on all seven. These observations were made after running the prespecified lower-tail analysis and therefore cannot redefine the primary estimand.

6.3 EXPLORATORY ACCESSIBLE-SUBSPACE PROFILES

Let U_α average the strongest α fraction, using the same $k = \lfloor \alpha q \rfloor$ and $\delta = \alpha q - k$ as Equation 6:

$$U_\alpha(\lambda) = \frac{\sum_{j=q-k+1}^{q} \lambda_{(j)} + \delta \lambda_{(q-k)}}{\alpha q}, \tag{15}$$

with the fractional term omitted when $\delta = 0$. Using clipped spectra, define

$$\text{U-SPA-20}(\theta, d, e) = 100 \, \frac{U_{.2}(\widehat{\lambda}_{\theta,d,e}) - \mathbb{E}_\pi U_{.2}(\widehat{\lambda}^{\pi}_{\theta,d,e})}{U_{.2}(\widehat{\lambda}^{\text{repeat}}_{d}) - \mathbb{E}_\pi U_{.2}(\widehat{\lambda}^{\text{repeat},\pi}_{d})}. \tag{16}$$

The null floor is model- and environment-specific; the clean split-taper denominator stays fixed. This post-hoc descriptor measures accessible upper-tail excess relative to the split-taper excess. It is not bounded to $[0, 100]$, a fraction of all retained information, or a validated selection rule.

For HMC, Motor-MV, MIMUL, ADFTD, and Workload, the clean mean and approximate conditional 95% half-width are 38.1 ± 16.2 (REVE), 30.9 ± 3.4 (EEGPT), 14.8 ± 4.3 (CBraMod), 13.2 ± 4.7 (CSBrain), and 5.1 ± 2.5 (LaBraM). The rule for this descriptive aggregate is $n_{\text{score}} \geq 13$, not an interval-width filter. Motor-MV remains included despite REVE's 14.0 ± 76.5 estimate. The archive does not establish preregistration of the threshold. Appendix E reports all seven cohorts, medians, and trimmed means. The intervals condition on fitted state maps, floors, and repeatability; they omit those sources of uncertainty and do not establish a ranking.

Across the same five cohorts, channel loss reduces the mean descriptor from 38.1 to 29.2 for REVE and from 30.9 to 18.5 for EEGPT. Taking the worst environment separately within each dataset gives means of 28.0 and 18.5. Effects are not uniformly adverse: some upper-tail scores increase under perturbation. These comparisons concern refitted probes and potentially different accessible subspaces; no perturbed-view jackknife was saved. Appendix E gives all 140 upper-tail scores and 35 within-dataset minima.

7 DISCUSSION

The audit finds concentrated access to a declared spectral state in pooled representations from five masked-reconstruction or masked-prediction models. The zero lower-tail estimates identify a failure to demonstrate broad, cross-subject linear recovery under this protocol. They do not prove that the encoder has destroyed the state or that downstream tasks must fail. At the population boundary $s = 0$, the lower-tail task bound offers no improvement over the zero-state baseline; tasks aligned with upper subspaces can still have a useful localized bound.

Where information may become inaccessible. The evaluated object includes the input adapter and epoch pooling. EEGPT averages time-local summary tokens; the other adapters average channel and temporal tokens. HMC's target has 29 fine windows plus one full-epoch block, so 96.7% of raw coordinate blocks are time-local, although their proportion in the retained PCA subspace is unknown. Pooling may remove local structure before probing. Position-aware tokens can still encode dynamics after averaging, so pooling does not make their recovery mathematically impossible. Token-level or window-aligned probes would separate this bottleneck from backbone loss; those experiments are not available here.

Grouped cross-validation also tests transfer between subjects. A state may remain decodable within each subject yet require different affine readouts across subjects. Subject derangement cannot distinguish this misalignment from within-subject information loss. Calibrated subject-specific probes would address that distinction but would answer a different, adaptation-dependent question. Nonlinear probes, larger cohorts, and alternative regularization remain similarly unresolved.

What the architecture comparison supports. LaBraM's discrete pretraining targets, REVE's continuous 4D positional encoding, and the other models' spatial and temporal designs are plausible sources of differing upper spectra (Jiang et al., 2024; El Ouahidi et al., 2025). LaBraM's tokenizer reconstructs both Fourier amplitude and phase; its frozen transformer output is continuous, so its low score is not evidence that a discrete embedding necessarily removed phase. REVE also uses

channelwise amplitude normalization in this adapter. Architecture, preprocessing, pooling, training data, and objectives vary together, preventing a causal attribution to any one choice. Reported reconstruction-related spectral biases (Kommineni et al., 2026a; Yu et al., 2026) motivate this audit, but do not explain these results by themselves. No conclusion about contrastive models follows without evaluating them.

Limits of the declared state. The assay retains selected second-order spectral directions, not every raw coordinate or every physiological property. It omits higher-order dynamics, cross-frequency coupling, and waveform morphology beyond the chosen windows. Narrow high-density bands rely on diagonal loading, and the archive cannot resolve how strongly retained PCA modes depend on that floor. Montages, recording protocols, and adapters further limit comparisons across datasets. A subject threshold alone does not ensure precision: control recovery and uncertainty must be checked in each cohort.

The useful output in this regime is the spectrum with its null and split-taper references. SPA-20 remains a stress test for broad coverage, while the upper-tail analysis describes what survives. Testing whether a prospectively specified summary predicts new spectrally mediated tasks would be a separate criterion-validity study (Zhang et al., 2025).

REPRODUCIBILITY STATEMENT

The implementation contains the spectral extractor, adapter code, grouped fitting, and control routines. Appendix C specifies the available protocol; Appendix A gives assumptions and proofs. The paper's consolidation script reads seven final JSON files and regenerates the reported environmental and uncertainty summaries without new encoder or probe runs. The archive preserves clipped model spectra, mean ordered null spectra, raw-spectrum diagnostics, and clean jackknife spectra. It does not include the state/embedding caches, fitted PCA loadings, complete raw spectra, or all individual control/null replicates. Thus certain requested sensitivity analyses cannot be reconstructed from this release.

ETHICS STATEMENT

All experiments use previously released datasets under their source terms. EEG representations can preserve subject or session identity (DelPozo-Banos et al., 2015); released artifacts should therefore follow source governance and should not treat pseudonymous identifiers as proof of anonymity. The method is an evaluation procedure and does not provide clinical diagnoses. Performance under the studied datasets and perturbations should not be interpreted as evidence of clinical safety or demographic fairness.

AI USE STATEMENT

Generative AI assisted with manuscript revision, mathematical exposition and checking, source-code auditing, and scripts that summarize existing experimental artifacts. No new experimental observations were generated by AI. The authors are responsible for verifying the mathematics, citations, numerical summaries, and final scientific claims.

REFERENCES

Diego Alvarez-Estevez and Roselyne Rijsman. Haaglanden Medisch Centrum Sleep Staging Database. *PhysioNet*, 2022. doi: 10.13026/t79q-fr32. Version 1.1.

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006. doi: 10.1002/mrm.20965.

Marcos DelPozo-Banos, Carlos M. Travieso, Christoph T. Weidemann, and Jesús B. Alonso. EEG biometric identification: a thorough exploration of the time-frequency domain. *Journal of Neural Engineering*, 12(5):056019, 2015. doi: 10.1088/1741-2560/12/5/056019.

Paolo Detti. Siena Scalp EEG Database. *PhysioNet*, 2020. doi: 10.13026/5d4a-j060. Version 1.0.0.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A Foundation Model for EEG—Adapting to Any Setup with Large-Scale Pretraining on 25,000 Subjects. In *Advances in Neural Information Processing Systems*, 2025.

Ary L. Goldberger, Luís A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):e215–e220, 2000. doi: 10.1161/01.CIR.101.23.e215.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked Autoencoders Are Scalable Vision Learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022.

Ji-Hoon Jeong, Kyung-Hwan Shim, Dong-Joo Kim, and Seong-Whan Lee. Multimodal Signal Dataset for 11 Intuitive Movement Tasks from Single Upper Extremity during Multiple Recording Sessions. *GigaScience*, 9(10):giaa098, 2020. doi: 10.1093/gigascience/giaa098.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI. In *International Conference on Learning Representations*, 2024.

Aditya Kommineni, Emily Zhou, Kleanthis Avramidis, Simon Bock Segaard, Jeppe Roden Münster, Andreas Peter Juhl Hansen, Takfarinas Medani, Tiantian Feng, Richard Leahy, and Shrikanth Narayanan. Aperiodic and Low-Frequency Spectral Bias in Reconstruction Based EEG Foundation Models. *arXiv preprint arXiv:2605.26434*, 2026a.

Aditya Kommineni, Emily Zhou, Kleanthis Avramidis, Tiantian Feng, and Shrikanth Narayanan. A Multi-dimensional Framework for Evaluating Generalization in EEG Foundation Models. In *Proceedings of the 7th Conference on Health, Inference, and Learning*, volume 333 of *Proceedings of Machine Learning Research*, pp. 221–244, 2026b.

Simon Kornblith, Jonathon Shlens, and Quoc V. Le. Do Better ImageNet Models Transfer Better? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2661–2671, 2019.

Demetres Kostas, Stéphane Aroca-Ouellette, and Frank Rudzicz. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, 15:653659, 2021. doi: 10.3389/fnhum.2021.653659.

Dingkun Liu, Yuheng Chen, Zhu Chen, Zhenyao Cui, Yaozhi Wen, Jiayu An, Jingwei Luo, and Dongrui Wu. EEG Foundation Models: Progresses, Benchmarking, and Open Problems. *arXiv preprint arXiv:2601.17883*, 2026.

Andreas Miltiadous, Konstantina D. Tzimourta, Nikolaos Giannakeas, Markos G. Tsipouras, Theodora Afrantou, Panagiotis Ioannidis, and Alexandros T. Tzallas. A Dataset of Scalp EEG Recordings of Alzheimer's Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG. *Data*, 8(6):95, 2023. doi: 10.3390/data8060095.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J. Miller, Gernot R. Müller-Putz, et al. Review of the BCI Competition IV. *Frontiers in Neuroscience*, 6:55, 2012.

David J. Thomson. Spectrum Estimation and Harmonic Analysis. *Proceedings of the IEEE*, 70(9): 1055–1096, 1982. doi: 10.1109/PROC.1982.12433.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. EEGPT: Pretrained Transformer for Universal and Reliable Representation of EEG Signals. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A Criss-Cross Brain Foundation Model for EEG Decoding. In *International Conference on Learning Representations*, 2025.

Jiamin Wu, Zichen Ren, Junyu Wang, Pengyu Zhu, Yonghao Song, Mianxin Liu, Qihao Zheng, Lei Bai, Wanli Ouyang, and Chunfeng Song. AdaBrain-Bench: Benchmarking Brain Foundation Models for Brain-Computer Interface Applications. *arXiv preprint arXiv:2507.09882*, 2025.

Wei Xiong, Jiangtong Li, Jie Li, and Kun Zhu. EEG-FM-Bench: A Comprehensive Benchmark for the Systematic Evaluation of EEG Foundation Models. *arXiv preprint arXiv:2508.17742*, 2025.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Biosignal Transformer for Cross-data Learning in the Wild. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Liuyin Yang, Qiang Sun, Ang Li, and Marc M. Van Hulle. Are EEG Foundation Models Worth It? Comparative Evaluation with Traditional Decoders in Diverse BCI Tasks. In *International Conference on Learning Representations*, 2026.

Junjie Yu, Zihan Deng, Jianyu Zhang, Junrong Mu, Jiahui An, Wenxiao Ma, Ziling Lu, Yue Wang, Yan Zhu, Kexin Lou, and Quanying Liu. Understanding and Correcting Low-Frequency Bias in EEG Foundation Model. *arXiv preprint arXiv:2608.01898*, 2026.

Weiyang Zhang, Xinyang Chen, Xiucheng Li, Kehai Chen, Weili Guan, and Liqiang Nie. Unified Transferability Metrics for Time Series Foundation Models. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Yuchen Zhou, Jiamin Wu, Zichen Ren, Zhouheng Yao, Weiheng Lu, Kunyu Peng, Qihao Zheng, Chunfeng Song, Wanli Ouyang, and Chao Gou. CSBrain: A Cross-scale Spatiotemporal Brain Foundation Model for EEG Decoding. In *Advances in Neural Information Processing Systems*, 2025.

Igor Zyma, Sergii Tukaev, Ivan Seleznov, Ken Kiyono, Anton Popov, Mariia Chernykh, and Oleksii Shpenkov. Electroencephalograms during mental arithmetic task performance. *Data*, 4(1):14, 2019. doi: 10.3390/data4010014.

A Proofs and additional theory

A.1 Formal setup and the linear population operator

Fix an environment e and condition on the declared target map. Let $\bar{S} \in \mathbb{R}^q$ and $Z_e \in \mathbb{R}^p$ be square-integrable, with $\mathbb{E}[\bar{S} \mid e] = 0$ and $\mathrm{Cov}(\bar{S} \mid e) = I_q$. The environment changes the input view, not the target distribution. Write $m_e = \mathbb{E}[\bar{S} \mid Z_e, e]$ and $A_e = \mathrm{Cov}(m_e \mid e)$. For centered $\widetilde{Z}_e = Z_e - \mathbb{E}[Z_e \mid e]$, define

$$m_e^{\mathrm{lin}} = \Sigma_{SZ,e}\Sigma_{ZZ,e}^{\dagger}\widetilde{Z}_e, \qquad A_e^{\mathrm{lin}} = \Sigma_{SZ,e}\Sigma_{ZZ,e}^{\dagger}\Sigma_{ZS,e}, \tag{17}$$

where † denotes the Moore–Penrose inverse. These are the best affine predictor and its covariance. The predictor is the L_2 projection onto the span of the representation coordinates and a constant, whether or not coefficients are unique. Projection orthogonality gives

$$\mathrm{Cov}(\bar{S} - m_e^{\mathrm{lin}} \mid e) = I_q - A_e^{\mathrm{lin}}, \qquad A_e - A_e^{\mathrm{lin}} = \mathrm{Cov}(m_e - m_e^{\mathrm{lin}} \mid e) \succeq 0. \tag{18}$$

Thus $0 \preceq A_e^{\mathrm{lin}} \preceq A_e \preceq I_q$. This linear operator is unregularized; a non-vanishing population ridge penalty defines a different residual-based functional.

A.2 Directional interpretation and endpoints

Proposition 1. For every unit $v \in \mathbb{R}^q$,

$$v^\top A_e v = \mathrm{Var}(\mathbb{E}[v^\top \bar{S} \mid Z_e, e] \mid e) = 1 - \mathbb{E}[\mathrm{Var}(v^\top \bar{S} \mid Z_e, e) \mid e]. \tag{19}$$

Consequently $0 \preceq A_e \preceq I_q$, and $A_e = I_q$ if and only if $\bar{S} = m_e$ almost surely. Independence implies $A_e = 0$, but the converse need not hold.

Proof. Apply total variance to $v^\top \bar{S}$, whose variance is one. If $A_e = I_q$, then $\mathbb{E}\|\bar{S} - m_e\|^2 = q - \mathrm{tr}(A_e) = 0$, proving perfect recovery; the converse is immediate. Independence makes m_e constant and centered. The zero operator means only that every conditional mean is constant: conditional variances or higher moments may still depend on Z_e. □

A.3 Data processing and representation coordinates

Proposition 2. Let $Z'_e = h(Z_e)$ for measurable deterministic h. Then $A_{Z',e} \preceq A_{Z,e}$, every ordered eigenvalue cannot increase, and neither can a lower- or upper-tail mean. If h has an almost-sure measurable inverse, equality holds.

Proof. Put $M = m_e(Z_e)$. The tower property gives $\mathbb{E}[\bar{S} \mid Z'_e, e] = \mathbb{E}[M \mid Z'_e, e]$. Total covariance yields

$$A_{Z,e} = A_{Z',e} + \mathbb{E}[\mathrm{Cov}(M \mid Z'_e, e) \mid e]. \tag{20}$$

The second term is positive semidefinite. Ordered-eigenvalue monotonicity follows from Courant–Fischer; tail means are nonnegative weighted sums of the ordered eigenvalues. An invertible measurable map preserves the generated sigma-algebra. □

Linear version and its limit. If $Z' = MZ + b$, its affine predictor space is a subspace of that of Z. Nested L_2 projections imply $A_{Z'}^{\mathrm{lin}} \preceq A_Z^{\mathrm{lin}}$, with equality for invertible M. This statement fails for general nonlinear h. For $Z \sim N(0,1)$ and $\bar{S} = (Z^2 - 1)/\sqrt{2}$, $A_Z^{\mathrm{lin}} = 0$, whereas $h(Z) = Z^2$ gives $A_{h(Z)}^{\mathrm{lin}} = 1$. The unrestricted operator is one in both cases.

Regularized implementation. For a fixed design matrix Z and target matrix Y, isotropic ridge minimizes $\|Y - ZW\|_F^2 + \eta\|W\|_F^2$. Replacing Z by ZQ for orthogonal Q and W by $Q^\top W$ preserves both terms at any $\eta \geq 0$. With the same folds, equivariant preprocessing, and consistent tie-breaking, cross-validation over the same grid also preserves predictions in exact arithmetic. A general invertible linear map changes the penalty geometry. Re-fitting per-coordinate variances after an orthogonal rotation is not equivariant, so StandardScaler followed by ridge need not be invariant. This is present in our pipeline; the rotation differences need not be numerical noise. Finite estimation error and hyperparameter selection also allow PCA projection to improve out-of-fold scores despite population data processing.

A.4 TARGET-COORDINATE INVARIANCE

Proposition 3. Let centered retained coordinates R have positive-definite covariance Σ. For invertible L, define $\bar{S} = \Sigma^{-1/2}R$ and $\bar{S}' = (L\Sigma L^\top)^{-1/2}LR$. Then $\bar{S}' = Q\bar{S}$ for orthogonal Q, and $A'_e = QA_eQ^\top$. The same similarity holds for A_e^{lin}.

Proof. Set $Q = (L\Sigma L^\top)^{-1/2}L\Sigma^{1/2}$. Multiplication gives $QQ^\top = I_q$. Both conditional expectation and affine least-squares projection commute with fixed linear output maps, giving the claimed covariance relation. □

This result fixes the retained space and uses its exact covariance. An invertible transformation of the raw features can change PCA selection, the reliability screen, or Ledoit–Wolf shrinkage. It therefore does not imply invariance of the entire finite state-construction procedure.

A.5 RISK FOR SPECTRALLY MEDIATED TASKS

Proposition 4. Suppose $Y = g_e(\bar{S}) + \varepsilon$, g_e is L_e-Lipschitz, $\mathbb{E}[\varepsilon^2] < \infty$, and $\mathbb{E}[\varepsilon \mid \bar{S}, Z_e, e] = 0$. For $s = \mathrm{SPA}_\alpha/100$ and $f_e(Z_e) = g_e(m_e(Z_e))$,

$$\mathbb{E}\|\bar{S} - m_e\|_2^2 = q - \mathrm{tr}(A_e) \le q(1-s), \tag{21}$$

$$\mathbb{E}[(Y - f_e)^2] - \mathbb{E}[\varepsilon^2] = \mathbb{E}[(g_e(\bar{S}) - g_e(m_e))^2] \le L_e^2\{q - \mathrm{tr}(A_e)\} \le L_e^2 q(1-s). \tag{22}$$

All expectations in this subsection are conditional on e. Replacing (m_e, A_e, s) with their linear population counterparts gives the same result.

Proof. Conditional-mean orthogonality gives the state-error identity. The mean eigenvalue is at least the mean of the lowest α fraction, hence $\mathrm{tr}(A_e)/q \ge L_\alpha(\lambda(A_e)) \ge s$. Lipschitz continuity bounds squared task discrepancy pointwise by $L_e^2\|\bar{S} - m_e\|^2$. The conditional zero-mean noise assumption removes the cross term when expanding $(Y - f_e)^2$. Affine projection orthogonality proves the linear version. □

The reference risk here is $\mathbb{E}[\varepsilon^2]$, attained by the state-informed predictor $g_e(\bar{S})$. It is not the risk of the optimal predictor using only Z_e. The plug-in $g_e(m_e)$ need not equal $\mathbb{E}[Y \mid Z_e, e]$ for nonlinear g_e, although the optimal Z_e predictor can only improve its risk.

Boundary behavior and dimensional scaling. When $s = 0$, predicting state zero gives the same generic bound $L_e^2 q$, because $\mathbb{E}\|\bar{S}\|^2 = q$. The tail-based guarantee then gives no improvement. Even a positive s may be uninformative relative to a particular task's variance: if $g(s) = v^\top s$ with $\|v\| = 1$, that variance is one, while $q(1-s) > 1$ whenever $s < 1 - 1/q$. The exact directional error is $v^\top(I_q - A_e)v$, not $q(1-s)$. The finite zero estimates in this paper neither estimate every task's error nor certify that population $s = 0$.

Localized bound. If $g_e(s) = h_e(Ps)$, $PP^\top = I_r$, and h_e is L_e-Lipschitz, then

$$\mathbb{E}[(h_e(P\bar{S}) - h_e(Pm_e))^2] \le L_e^2\mathbb{E}\|P(\bar{S} - m_e)\|^2 = L_e^2\,\mathrm{tr}\big(P(I_q - A_e)P^\top\big). \tag{23}$$

For rows of P spanning the top r eigenvectors, the right-hand side is $L_e^2\sum_{j=q-r+1}^{q}(1 - \lambda_{e,(j)})$. This improves on the r-dimensional zero-state bound when the task depends on that accessible subspace. It is conditional on task alignment, not a guarantee that every task uses the best subspace. A fixed task projection may also align differently across environments. The linear version follows identically.

A.6 WHY THE DECLARED STATE CANNOT BE REMOVED

Let $X = (U, V)$ for independent non-degenerate bits, $Z_1 = U$, and $Z_2 = V$. Representation Z_1 perfectly recovers state U and is uninformative about V; Z_2 reverses the ordering. A representation ordering therefore requires a declared state or task family.

A.7 FRACTIONAL TAIL FUNCTIONALS

Equation 6 is the integral of the empirical quantile function over $[0, \alpha]$, divided by α. Equation 15 integrates over $[1-\alpha, 1]$. Both are continuous in $\alpha > 0$ and 1-Lipschitz in the ordered eigenvalues

under the sup norm. When αq is an integer, the fractional term is absent; when $\alpha q < 1$, each functional equals its respective extreme eigenvalue. In particular, $\alpha = 1$ needs neither λ_{q+1} nor λ_0.

A.8 Population operator versus residual plug-in estimate

At the unrestricted or best affine population predictor, the residual has mean zero and is orthogonal to the prediction, so $A_e = I_q - \mathrm{Cov}(\bar{S} - m_e \mid e)$ (and analogously for A_e^{lin}). A finite cross-fitted predictor need not satisfy this orthogonality. Nevertheless,

$$\widehat{\Sigma}_{\mathrm{res},e} \succeq 0 \quad \Longrightarrow \quad \widehat{\Sigma}_S - \widehat{\Sigma}_{\mathrm{res},e} \preceq \widehat{\Sigma}_S \quad \Longrightarrow \quad \widehat{B}_e \preceq I_q. \tag{24}$$

Thus raw eigenvalues cannot exceed one in exact arithmetic, provided the same positive-definite target covariance is used. They can be arbitrarily negative. Upper clipping is a numerical safeguard; lower clipping can change ordering and create exact ties at zero. A clipped zero is not a confidence statement about population access.

The implementation uses centered sample covariances. For any finite predictor, the population analogue satisfies

$$\mathbb{E}\|\bar{S} - \widehat{m}_e\|^2 = \mathrm{tr}\,\mathrm{Cov}(\bar{S} - \widehat{m}_e) + \|\mathbb{E}[\bar{S} - \widehat{m}_e]\|^2. \tag{25}$$

The covariance score does not penalize the second term. Equation 10 cannot therefore be applied to a finite covariance score without controlling bias and estimation error.

A.9 What a consistent linear probe would estimate

A sufficient regime fixes q, embedding dimension, and $|\mathcal{E}|$; lets the numbers of independent state-fit and scoring subject clusters grow; assumes bounded cluster sizes and fourth moments; separates population reliability values from their threshold and the retained PCA eigenspace from excluded eigenvalues; and bounds retained target covariance eigenvalues away from zero. The limiting state is the declared, finite-window, regularized CSD feature map. Consistency for a latent physiological spectrum would need additional spectral-estimation assumptions, not merely more subjects.

Suppose fitted state maps converge (up to orthogonal coordinates) and held-out affine predictions satisfy $\|\widehat{m}_e - m_e^{\mathrm{lin}}\|_{L_2} \to 0$ on independent subject clusters. Then cluster laws of large numbers and convergence of predictions give

$$\widehat{\Sigma}_S \to I_q, \qquad \widehat{\Sigma}_{\mathrm{res},e} \to I_q - A_e^{\mathrm{lin}}, \qquad \widehat{\text{L-SPA}}_\alpha \to 100 \min_e L_\alpha(\lambda(A_e^{\mathrm{lin}})). \tag{26}$$

If the limiting state map is not exactly whitening, score-side covariance normalization instead yields the equivalent rewhitened operator. Continuity of inverse square roots, eigenvalues, clipping, tail means, and the finite minimum proves the result. For fixed-dimensional ridge with stable feature scaling and nonsingular feature covariance, an effective penalty tending to zero and growing inner/outer training samples are sufficient for convergence to the affine predictor. In the sum-of-squares convention, this concerns η/n, not η alone. Fixed nonzero effective shrinkage generally has a different limit. Convergence to unrestricted SPA additionally requires $m_e^{\mathrm{lin}} = m_e$, or an estimator consistent for the unrestricted conditional mean. None of these asymptotic assumptions is a finite-sample guarantee for this audit.

B Algorithm

1. **Declare the assay.** Fix the montage, preprocessing, spectral bands, time resolutions, environments $\mathcal{E}$, tail mass α, reliability threshold, maximum state rank, and decoder family before checkpoint comparison.
2. **Split subjects.** Partition subjects into state-fit and score sets. Keep every epoch and every transformed view of one subject on the same side.
3. **Build three raw states.** For each clean state-fit epoch, compute full-taper and disjoint half-taper CSD features.
4. **Fit the reliable state map.** Fit PCA to the mean of the halves, retain directions whose half-to-half correlation exceeds the declared threshold on state-fit data, and fit centering and whitening on full-taper coordinates.

5. **Freeze the target.** Apply the fitted map to clean score epochs. Never recompute the state basis, reliability selection, or target from a perturbed checkpoint view.
6. **Extract checkpoint views.** For each $e \in \mathcal{E}$, compute the declared adapted, perturbed, and pooled checkpoint view $Z_{\theta,e}$ from the frozen released checkpoint.
7. **Cross-fit.** In each environment, use nested subject-grouped folds to select the decoder hyperparameters and obtain predictions only for outer-held-out subjects.
8. **Estimate the spectrum.** Form Equation 14, retain raw eigenvalues, and clip to $[0, 1]$ for the finite summary.
9. **Run controls through the same estimator.** Evaluate the oracle, subject derangements, coordinate rotation, lossy projection, and split-taper repeatability.
10. **Report.** Report clipped spectra, available raw diagnostics, per-environment scores, null floors, split-taper references, subject counts, and sensitivity summaries. Future runs should also retain complete raw spectra and all control replicates.

C Datasets, preprocessing, and adapters

Table 2: Dataset pools and canonical target epochs. Channel counts refer to the target route, not necessarily every checkpoint input. Subject counts describe the benchmark pool before state-fit/score caps.

Dataset	paradigm	epoch	channels	subjects
BCIC-IV-2a (Tangermann et al., 2012)	motor imagery	4 s	22	9
HMC (Alvarez-Estevez & Rijsman, 2022)	sleep staging	30 s	4	151
ADFTD (Miltiadous et al., 2023)	dementia	10 s	19	88
Motor-MV (Goldberger et al., 2000)	motor execution/imagery	4 s	60	110
Siena (Detti, 2020)	seizure detection	10 s	27	14
Workload (Zyma et al., 2019)	cognitive workload	4 s	19	36
MIMUL (Jeong et al., 2020)	upper-limb movement	5 s	58	25

C.1 Sampling and target-input provenance

The table's channel counts describe the *target route*. The driver uses clean tensors from the EEGPT data adapter as the canonical state input, after its channel selection and fixed 10^{-3} amplitude scaling. Thus the target montage contains 60 of Motor-MV's 64 listed channels, 58 of MIMUL's 60, and 27 of Siena's 29. Other checkpoint adapters can receive additional channels (Table 4); identical sensor sets across models should not be inferred from Table 2. HMC's four signals remain referenced derivations even though their names are reduced to F4, C4, O2, and C3 for model lookup. The Workload source is the 36-subject EEGMAT mental-arithmetic dataset (Zyma et al., 2019); it is not STEW.

The dataset builders filter and resample to 200 Hz before the model adapters. The base configuration uses a 0.1-Hz high-pass and a 100-Hz low-pass when the original sampling rate permits it; BCIC-IV-2a declares 0.5–45 Hz. Notch handling follows dataset configuration (50 Hz, or 60 Hz for Motor-MV, when not already marked notched). The builder applies band-pass filtering only when recording length exceeds its required filter length. These are source-defined rules, not reconstructed preprocessing logs for each recording.

The driver pools original train/validation/test subject lists, sorts identifiers by length and then lexicographically, assigns the first half to state fitting and the second to scoring, and caps each at 30 subjects. Within each original split, up to 40 evenly spaced rows are selected per subject. The original splits are subject-disjoint; Table 1 records the resulting counts. No label enters state fitting or ridge tuning. A fixed state basis is reused for all models and environments. The penalty grid is $\{10^{-3+7j/19} : j = 0, \ldots, 19\}$, with up to five outer and four inner GroupKFold folds and one shared multi-output penalty selected by aggregate MSE.

C.2 MODEL ADAPTERS AND THE POOLED REPRESENTATION

The shared bridge hooks the tensor supplied to the classifier, before its learned task head. For a tensor of shape $[B, T, C, d]$ it averages axes T, C; for $[B, T, d]$ it averages T. All five routes below produce the first form. The classifier's output is not the representation. Table 3 documents the source implementation; Table 4 crosses its rules with the seven dataset montages and durations.

Table 3: Source-defined adapters. All operate on the benchmark's 200-Hz stream. No additional native-rate conversion, spatial spline interpolation, or zero-filled full montage is implemented in these routes. The declared patch settings are from the released configuration files.

Model	channel and amplitude handling	duration and representation handling
EEGPT	Select names in the standard supported list; scale by 10^{-3}. Channel convolution is disabled in the released configurations.	64-sample convolutional windows, stride 32, retaining complete windows. Mean over time and four learned spatial summary tokens; $d = 512$.
REVE	Select supported names from the restricted electrode bank; use its positions. Z-score each channel within the epoch.	200-sample windows, stride 180; drop an incomplete trailing window. Mean over channel/time tokens; $d = 512$. Its attention-pooling method is not called.
CSBrain	Select names in the region map; scale by 10^{-2}; sort by brain region and within-region topology.	Truncate to complete 200-sample patches; mean over channel/time output tokens; $d = 200$.
CBraMod	Retain the input montage and scale by 10^{-2}.	Reshape into 200-sample patches (all audited durations divide exactly); mean over channel/time tokens; $d = 200$.
LaBraM	Select names from its channel vocabulary; scale by 10^{-2}.	Complete 200-sample patches; time embeddings are linearly interpolated when needed. Return normalized patch tokens, excluding the class token, and average channel/time axes; $d = 200$.

Table 4: Static reconstruction from the archived dataset and adapter source, not a recovered runtime manifest. Counts are selected input channels for each checkpoint. Last column: time windows for EEGPT / REVE / each of CSBrain, CBraMod, and LaBraM. No sliding chunk evaluation or epoch padding is used.

Dataset	samples	EEGPT	REVE	CSBrain	CBraMod	LaBraM	windows
HMC	6000	4	4	4	4	4	186/33/30
Motor-MV	800	60	64	64	64	64	24/4/4
MIMUL	1000	58	60	60	60	60	30/5/5
ADFTD	2000	19	19	19	19	19	61/11/10
Workload	800	19	19	19	19	19	24/4/4
Siena	2000	27	29	29	29	29	61/11/10
BCIC-IV-2a	800	22	22	22	22	22	24/4/4

The shared selector retains supported data rows in their original montage order and supplies model channel IDs from the supported-name list; CSBrain then applies its regional ordering. HMC strips reference suffixes for lookup; ADFTD, Workload, and Siena map T3/T4/T5/T6 to T7/T8/P7/P8; Motor-MV strips punctuation; BCIC-IV-2a uses its explicit 22-electrode map. Unsupported electrodes are dropped, not synthesized. The source-derived channel lists and omissions for all 35 combinations are in `results/adapter_routes.json`. Historical per-epoch channel manifests and caches are absent, so exact run-time ordering and position alignment cannot be independently certified from final spectra. These routing and normalization choices belong to the evaluated representation and may affect spectral access.

C.3 BAND RANK, LOADING, AND THE LOG MAP

For a band with B_b discrete Fourier bins, an unregularized K-taper CSD is a weighted sum of KB_b complex rank-one outer products. Hence its rank is at most $\min(C, KB_b)$; $KB_b \geq C$ is necessary

but not sufficient for full rank. Bands use left-closed, right-open intervals. At two seconds, bin spacing is 0.5 Hz, giving seven delta bins in $[0.5, 4)$ and eight theta bins in $[4, 8)$. Their full-taper rank bounds are therefore 42 and 48, and half-taper bounds are 21 and 24. For 58–60 target channels both narrow-band full-taper matrices are necessarily singular before loading.

The code adds 10^{-7} times mean channel power to the diagonal, with a 10^{-15} absolute safeguard for zero power. It then normalizes by trace and floors shape eigenvalues at $10^{-7}/C$ before the Hermitian logarithm (again with a numerical safeguard). High-density narrow bands therefore operate in a *loaded full-rank regime*. Null-space eigenvalues are near the numerical floor; their eigenvectors can still depend on the signal. The log map retains this dependence, so floor sensitivity cannot be dismissed as an irrelevant constant. State-fit split-taper screening is not a substitute for a loading-sensitivity analysis. The final JSON files do not preserve fitted PCA loadings or raw state caches, so neither mode loading on floor-controlled directions nor an alternative-loading result is available.

C.4 Input environments and null construction

Perturbations are applied to each model's adapted batch tensor, after its dataset normalization and before the encoder; the clean canonical target is unchanged. Channel loss samples $\lfloor C/4 \rfloor$ channels without replacement per epoch and sets them to zero. Thus the exact fraction depends on the adapter channel count. Resampling keeps every second sample and linearly interpolates to the original length, using endpoint extension for the final sample; this perturbation does not insert a separate antialiasing filter. Colored noise uses independent complex Gaussian coefficients per channel multiplied by $1/f$ in amplitude (approximately $1/f^2$ power), inverse-transforms them, and scales their standard deviation to $10^{-10/20}$ times the signal standard deviation. This corresponds to 10 dB signal-to-noise power ratio on centered signals.

The perturbation seed derives from a hash of environment and epoch index. Python's string hash is process-dependent unless `PYTHONHASHSEED` is fixed, and the final JSONs do not record that setting or the realized perturbations. Exact regeneration of these views therefore requires the original process setting or cached tensors; within one process the same index and environment reproduce the transform.

A null pairs each recipient subject with a different donor. Donor epochs are sampled without replacement when sufficient; otherwise all donor epochs are used and the shortfall is sampled with replacement. This is a subject-block mismatch reference, not a proof of exchangeability or an exact permutation test under unequal subject sizes. Each model/view uses 20 such fits, while the split-taper denominator uses five. The split-taper control predicts half A from half B after both pass through the fixed full-taper state map; it computes its own target covariance for half A.

D Complete lower-tail and control results

Table 5: Clean-input controls: lower-tail / mean / upper-tail accessible variance. The checkpoint is EEGPT. Rotation precedes coordinatewise scaling; PCA32 is fitted on all score embeddings before probe cross-validation and is therefore transductive. These controls diagnose the finite estimator, not population invariance or monotonicity.

Dataset	oracle	split-taper	checkpoint	rotation	PCA32
HMC	1.000/1.000/1.000	.852/.945/.996	.000/.208/.831	.000/.206/.827	.000/.170/.756
Motor-MV	1.000/1.000/1.000	.843/.945/.998	.000/.086/.414	.000/.082/.397	.000/.079/.359
MIMUL	1.000/1.000/1.000	.780/.923/.997	.000/.117/.500	.000/.119/.506	.000/.104/.464
ADFTD	1.000/1.000/1.000	.736/.889/.991	.000/.059/.287	.000/.060/.291	.000/.069/.323
Workload	1.000/1.000/1.000	.424/.781/.992	.000/.057/.276	.000/.057/.276	.000/.041/.183
Siena	1.000/1.000/1.000	.354/.804/.995	.000/.064/.302	.000/.063/.298	.000/.086/.343
BCIC-IV-2a	1.000/1.000/1.000	.019/.579/.945	.000/.050/.230	.000/.055/.253	.000/.062/.248

All 140 environment-specific lower-tail values are zero. In the clean condition, 34 of 35 model–dataset medians are zero. Counts above declared thresholds and raw negative eigenvalue counts are in `results/spectrum_summary.csv`. The JSONs retain clipped model spectra and mean ordered null spectra, but only raw minimum, mean, and negative-count diagnostics; individual raw

spectra and null replicates are absent. PCA32 raises the EEGPT upper tail from .287 to .323 on ADFTD, .302 to .343 on Siena, and .230 to .248 on BCIC-IV-2a. All lower tails remain zero, so those ties supply no evidence for data-processing monotonicity.

E EXPLORATORY UPPER-TAIL RESULTS

E.1 CLEAN-INPUT VALUES AND CONDITIONAL UNCERTAINTY

Table 6: Clean U-SPA-20: point estimate ± approximate conditional jackknife 95% half-width. The probe is refitted after each subject deletion; the state map, model-specific null floor, and repeatability denominator remain fixed. The intervals omit uncertainty in those quantities. All values and summaries are computed from unrounded artifacts before formatting.

Model	HMC	Motor-MV	MIMUL	ADFTD	Workload	Siena	BCIC-IV-2a
REVE	65.3±2.1	14.0±76.5	53.9±8.4	48.4±24.9	8.7±4.0	34.7±13.1	26.9±23.7
EEGPT	78.1±6.7	13.8±5.7	39.0±13.5	12.8±4.8	10.5±2.2	17.6±22.7	10.1±22.6
CBraMod	35.9±7.4	4.4±8.6	23.0±17.7	9.0±3.5	1.5±1.4	12.2±8.1	4.3±11.2
CSBrain	46.6±8.3	5.2±3.0	5.4±19.5	7.5±9.8	1.4±2.0	13.8±17.1	5.8±16.7
LaBraM	10.8±4.3	2.6±2.2	5.4±11.1	6.1±1.7	0.8±1.2	6.0±10.8	−1.8±3.1

For dataset d and model θ, let $u_{d,(-i)}$ be the raw upper-tail mean after deleting subject i and rerunning nested ridge, and let D_d denote the split-taper excess denominator in Equation 16. With n_d score subjects, the reported conditional standard error is

$$\widehat{\mathrm{SE}}_d = \frac{100}{D_d}\left[\frac{n_d-1}{n_d}\sum_{i=1}^{n_d}(u_{d,(-i)}-\bar{u}_{d,(-\cdot)})^2\right]^{1/2}. \tag{27}$$

Half-widths are $1.96\widehat{\mathrm{SE}}_d$. They are approximate sensitivity summaries, particularly with few subjects and cross-validation model selection. They do not propagate baseline or state-estimation variability. Omitted quantities can be correlated with the numerator, so these conditional errors are not mathematically guaranteed lower bounds on total standard errors. Joint resampling is needed for full ratio uncertainty and is not reconstructible from the retained summaries.

Table 7: Sensitivity of the five-cohort clean aggregate. Eligibility uses $n_{\mathrm{score}} \geq 13$: HMC, Motor-MV, MIMUL, ADFTD, and Workload. Mean intervals and SEs condition on baseline quantities and assume independent datasets. The trimmed mean removes the largest and smallest cohort value, leaving three. These are descriptive alternatives, not additional selection criteria.

Model	mean ± 95% half-width	conditional SE	median	trimmed mean
REVE	38.1 ± 16.2	8.27	48.4	38.8
EEGPT	30.9 ± 3.4	1.73	13.8	21.9
CBraMod	14.8 ± 4.3	2.18	9.0	12.1
CSBrain	13.2 ± 4.7	2.41	5.4	6.0
LaBraM	5.1 ± 2.5	1.25	5.4	4.7

For the fixed set $\mathcal{D}_5$ of five eligible cohorts, the equal-weight mean is $\bar{U} = 5^{-1}\sum_d U_d$. Its conditional error under dataset independence is $\widehat{\mathrm{SE}}(\bar{U}) = \sqrt{\sum_d \widehat{\mathrm{SE}}_d^2}/5$. This describes variation within these datasets, not generalization to a population of datasets. The sample-size filter is explicit and independent of interval width, but the archive supplies no evidence it was preregistered. Retaining Motor-MV does not make its high-variance REVE estimate precise; the resulting uncertainty is visible in the aggregate. Siena and BCIC-IV-2a remain in all per-dataset tables.

Negative U-SPA-20 values mean the observed upper tail lies below its estimated floor; underlying clipped eigenvalues remain in $[0, 1]$. Raw split-taper upper-tail means range from .945 to .998, while the *floor-subtracted denominators* range from .856 to .960. These two quantities must not be conflated. A large upper-tail denominator does not certify lower-tail recoverability.

E.2 ALL INPUT ENVIRONMENTS

We use the saved observed spectrum and the mean of the 20 *ordered* null spectra for each environment. Since U_α is a fixed weighted sum of ordered eigenvalues, applying it to that mean equals averaging the 20 individual upper-tail values. This identity does not involve eigenvalues of an averaged operator. It reconstructs the environmental null floors exactly from the available artifacts.

Table 8: Exploratory U-SPA-20 across all four environments. "Worst" is the minimum of the four scores within the given dataset/model; the last column identifies that environment. Each cell uses its own null floor and the same clean split-taper denominator. No perturbed-environment jackknife intervals are available.

Dataset	model	clean	channel loss	resampling	noise	worst	environment
HMC	REVE	65.3	50.8	66.0	63.1	50.8	channel loss
HMC	EEGPT	78.1	50.4	77.4	74.7	50.4	channel loss
HMC	CBraMod	35.9	24.8	38.7	29.4	24.8	channel loss
HMC	CSBrain	46.6	30.6	46.1	37.9	30.6	channel loss
HMC	LaBraM	10.8	9.9	19.4	10.6	9.9	channel loss
Motor-MV	REVE	14.0	13.0	15.6	14.1	13.0	channel loss
Motor-MV	EEGPT	13.8	9.3	13.5	13.8	9.3	channel loss
Motor-MV	CBraMod	4.4	5.8	4.4	4.5	4.4	resampling
Motor-MV	CSBrain	5.2	4.9	5.1	5.0	4.9	channel loss
Motor-MV	LaBraM	2.6	2.4	2.3	2.6	2.3	resampling
MIMUL	REVE	53.9	43.8	52.5	52.0	43.8	channel loss
MIMUL	EEGPT	39.0	15.4	39.5	40.2	15.4	channel loss
MIMUL	CBraMod	23.0	20.1	20.2	22.0	20.1	channel loss
MIMUL	CSBrain	5.4	18.6	8.2	16.3	5.4	clean
MIMUL	LaBraM	5.4	4.8	1.7	6.4	1.7	resampling
ADFTD	REVE	48.4	30.8	45.8	24.9	24.9	colored noise
ADFTD	EEGPT	12.8	9.3	13.0	12.3	9.3	channel loss
ADFTD	CBraMod	9.0	8.2	9.8	8.7	8.2	channel loss
ADFTD	CSBrain	7.5	6.6	7.0	7.3	6.6	channel loss
ADFTD	LaBraM	6.1	4.6	2.6	6.1	2.6	resampling
Workload	REVE	8.7	7.5	8.8	8.1	7.5	channel loss
Workload	EEGPT	10.5	7.9	10.6	10.4	7.9	channel loss
Workload	CBraMod	1.5	1.7	1.6	1.1	1.1	colored noise
Workload	CSBrain	1.4	1.0	1.4	1.4	1.0	channel loss
Workload	LaBraM	0.8	0.5	0.7	0.9	0.5	channel loss
Siena	REVE	34.7	30.1	37.7	30.8	30.1	channel loss
Siena	EEGPT	17.6	14.6	21.5	17.2	14.6	channel loss
Siena	CBraMod	12.2	7.4	22.7	12.2	7.4	channel loss
Siena	CSBrain	13.8	12.5	15.5	13.4	12.5	channel loss
Siena	LaBraM	6.0	5.6	6.8	5.8	5.6	channel loss
BCIC-IV-2a	REVE	26.9	27.2	29.8	24.8	24.8	colored noise
BCIC-IV-2a	EEGPT	10.1	4.2	8.1	8.1	4.2	channel loss
BCIC-IV-2a	CBraMod	4.3	2.1	1.0	2.0	1.0	resampling
BCIC-IV-2a	CSBrain	5.8	10.7	5.6	5.5	5.5	colored noise
BCIC-IV-2a	LaBraM	-1.8	-2.1	-0.0	-1.9	-2.1	channel loss

Table 9: Equal-weight environmental means over the same five eligible cohorts. The final column averages within-dataset minima, $5^{-1}\sum_{d\in\mathcal{D}_5}\min_e U_{d,e}$; it is not the minimum of the four column means.

Model	clean	channel loss	resampling	noise	mean of worst
REVE	38.1	29.2	37.7	32.5	28.0
EEGPT	30.9	18.5	30.8	30.3	18.5
CBraMod	14.8	12.1	14.9	13.1	11.7
CSBrain	13.2	12.3	13.6	13.6	9.7
LaBraM	5.1	4.5	5.3	5.3	3.4

Channel loss has the largest mean reduction for REVE and EEGPT, but the worst view depends on the dataset and checkpoint. REVE/ADFTD drops from 48.4 clean to 24.9 with noise, while CSBrain/MIMUL rises from 5.4 clean to 18.6 under channel loss. Such changes are compatible with a finite refitted probe and with input views that are not ordered by representation data processing. Sorted upper eigenvalues can also describe different directions in different environments. These tables

establish descriptive sensitivity, not significance, retention of one common subspace, or performance on unseen perturbations.

F CANDIDATE SCALAR DIAGNOSTICS

Nine clean-input summaries were inspected only after L-SPA-20 collapsed. The lower 20% and lower 50% were degenerate on every dataset. The clipped mean was non-degenerate but had negative clipped-versus-raw Kendall correlation on six datasets, indicating that model order depended on eigenvalue clipping. Upper-tail summaries were non-degenerate and rank-stable to clipping, but narrower tails increasingly emphasize only a few directions. The largest-20% mean was the deepest upper-tail window that consistently cleared its own derangement floor, motivating Equation 16 as a descriptive follow-up. This selection is post hoc; confirmatory use requires new checkpoints or datasets and a preregistered rule.

G RELATIONSHIP TO OTHER EVALUATION VIEWS

Frozen task probes, fine-tuning, robustness tests, identity decodability, and spectral accessibility answer distinct questions. The broader project artifacts contain these views, but they are not averaged into SPA-20. In particular, a downstream label can depend on a small accessible subspace, while SPA-20 penalizes any large blind spot in the declared state. Conversely, high spectral accessibility need not imply usefulness for a label unrelated to second-order spectra. Reporting disagreement is more informative than constructing a composite whose weighting has no scientific interpretation.