RHYTHM FLOW: BRIDGING THE GAP BETWEEN FUNCTIONAL FLUCTUATIONS AND NEURAL SPIKES THROUGH OSCILLATORY SYNCHRONIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Understanding the complex relationship between spatially coarse millisecond-scale neural oscillations and fine-resolution yet slow hemodynamic fluctuations remains a fundamental challenge in multimodal neuroimaging. Although recent generative models attempt to synthesize functional magnetic resonance imaging (fMRI) from concurrent electroencephalography (EEG), current 'black-box' architectures fail to model intrinsic neurovascular coupling mechanisms, yielding anatomically plausible but functionally incoherent fMRI data. To bridge this gap, we introduce *Rhythm Flow*, a biologically-principled generative framework in which one oscillatory-synchronization mechanism governs both the EEG encoder and the residual dynamics it drives. Our approach decomposes the generated $3D + t$ latent space into three components: a time-invariant anatomy term, a low-rank EEG residual produced by a rhythm-gated spiking encoder (named as *VibeSync*), and a stochastic correction flow. The correction flow uses the same oscillator-and-spiking computation (*VibeSync*) to capture population-level dynamics that the EEG residual leaves unexplained. To test whether the predictions depend on the specific EEG input, we swap the input EEG, which changes only the EEG residual, and check without retraining whether temporally matched EEG explains the observed fMRI better than temporally shifted or cross-subject controls. Across two simultaneously recorded EEG–fMRI cohorts, temporally matched EEG outperforms both controls, and without the correction flow, the EEG residual is positive for nine of ten held-out subjects under five-fold subject cross-validation (one-sided Wilcoxon $p = 0.024$), indicating that the EEG specificity generalizes to unseen subjects. On 1,152 subjects of the Healthy Brain Network, *Rhythm Flow* further outperforms all compared models in both representational similarity (RSA-ROI) and voxel fidelity. Beyond empirical gains, *Rhythm Flow* turns EEG reliance into a quantity measured directly on a trained model, making the claim that a generative EEG-to-fMRI model has learned physiological coupling falsifiable.

1 INTRODUCTION

The human brain is a complex, dynamic system whose activity can be observed through complementary windows. Functional magnetic resonance imaging (fMRI) provides spatially resolved whole-brain measurements through the blood-oxygen-level-dependent (BOLD) response, but its temporal resolution is limited and its acquisition requires expensive, immobile scanners (Glover, 2011). In contrast, electroencephalography (EEG) records electrophysiological activity at millisecond resolution with portable equipment, yet localizing its sources from scalp measurements is fundamentally ill-posed (Michel & Brunet, 2019). Since the two modalities offer complementary views of the same neural dynamics, generating an individualized 4-D fMRI sequence from EEG, conditioned on the subject's structural MRI (sMRI), provides a computational bridge between them. In this context, our work asks whether accessible electrophysiology can be projected into a spatial representation constrained by known anatomy.

This task presents three challenges. (1) BOLD is an indirect and temporally smoothed consequence of neural activity, so EEG does not map uniquely to fMRI intensity. (2) The output is a high-

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.

dimensional spatiotemporal field, which requires modeling both volumetric organization and frame-to-frame dynamics. (3) Functional organization varies substantially across individuals (Mueller et al., 2013; Gordon et al., 2017), which requires separating the dynamics carried by EEG from the anatomical support supplied by sMRI. Prior studies established that the two modalities share informative structure, from relating EEG frequency bands to fMRI connectomes (Deligianni et al., 2014) to learned generators for voxel-level (Calhas & Henriques, 2023) and region-level (Li et al., 2024) BOLD prediction. However, these methods typically produce template-space volumes or regional time series, and lack a system-level mechanism that ties EEG rhythms to the generated dynamics.

Following the notion that oscillatory synchronization coordinates activity among distributed neural populations (Varela et al., 2001; Dan & Wu, 2026), we conceptualize each EEG electrode as a coupled oscillator whose phase evolution carries the rhythm-governed state for fMRI generation. To that end, we develop *VibeSync*, which combines graph-coupled phase dynamics with a rhythm-gated spiking pathway (Yeung & Strogatz, 1999; Neftci et al., 2019) and aggregates the resulting membrane trajectories into conditioning tokens. However, injecting these tokens into a single flow-matching model (Lipman et al., 2023) conditioned on both EEG and sMRI creates an identifiability problem. Since the network is free to lean on the anatomical prior from sMRI, nothing in its architecture or objective compels it to exploit EEG, and whether it does can only be checked indirectly by retraining an EEG-ablated variant.

In this work, we bridge the gap by construction. Our *Rhythm Flow* factorizes the 4-D fMRI latent into three additive, role-restricted terms: (1) a time-invariant anatomy branch driven only by sMRI, (2) a deterministic low-rank residual driven only by EEG, and (3) a stochastic correction flow blind to both modalities, whose stop-gradient coordinate prevents it from absorbing the EEG branch. The major contributions of our work are three-fold. (1) We develop *VibeSync*, a rhythm-gated EEG encoder that couples spherical phase dynamics with spiking integration to produce structured conditioning tokens. (2) We identify the identifiability failure of jointly conditioned flow models and introduce *Rhythm Flow*, which guarantees an isolated EEG pathway by construction. (3) We devise a retraining-free paired evaluation against temporally shifted and cross-subject EEG, which also serves as a training signal for EEG specificity. On simultaneous EEG–fMRI data (Lioi et al., 2020), temporally matched EEG outperforms both controls, and without the correction flow, the EEG residual is positive for nine of ten held-out subjects under subject cross-validation. On the 1152-subject Healthy Brain Network (HBN) cohort (Alexander et al., 2017), *Rhythm Flow* further outperforms single-flow generators as well as template and region-level baselines in both representational similarity and voxel fidelity.

2 METHOD

Rhythm Flow factorizes the 4-D fMRI latent into three additive, role-restricted terms (Figure 1),

$$\widehat{\mathbf{Z}}_1 = \mathbf{A}(\mathbf{M}) + \mathbf{D}(\mathbf{E}) + \mathbf{C}_\theta, \qquad \mathrm{mean}_T(\mathbf{D}) = \mathrm{mean}_T(\mathbf{C}_\theta) = \mathbf{0}, \tag{1}$$

where $\mathbf{A}$ is a *time-invariant* anatomy branch conditioned only on the sMRI $\mathbf{M}$, $\mathbf{D}$ is a *deterministic* EEG residual conditioned only on the EEG $\mathbf{E}$, and $\mathbf{C}_\theta$ is a *stochastic* correction flow that observes neither modality. The architecture thus fixes the role of each term. Because $\mathbf{A}$ is constant over time, all temporal fluctuations originate from $\mathbf{D}$ or $\mathbf{C}_\theta$, and because both have zero temporal mean, EEG cannot alter the static baseline. Since $\mathbf{D}$ is an explicit function of EEG alone, its dependence on EEG can be measured on a trained model without retraining (Section 2.4). The pretrained variational autoencoder (VAE) is frozen, and $\mathbf{A}$, $\mathbf{D}$, and $\mathbf{C}_\theta$ are optimized jointly.

2.1 PROBLEM FORMULATION AND STATIC ANATOMY BRANCH

Each sample consists of an EEG segment $\mathbf{E} \in \mathbb{R}^{N\times L}$ with N electrodes and L time samples, the subject's T1-weighted sMRI $\mathbf{M}$, and a window of T fMRI frames $\mathbf{Y} \in \mathbb{R}^{T\times H_0\times W_0\times D_0}$ (EEG length and repetition time (TR) per cohort in Appendix C.1). In simultaneous recordings, acquisition triggers align $\mathbf{E}$ with $\mathbf{Y}$, so the matched EEG of a window is physically distinguishable from the EEG of any other window. A frozen 3-D VAE with encoder $\mathcal{E}_\phi$ and decoder $\mathcal{G}_\phi$ encodes the sMRI and each fMRI frame as

$$\mathbf{Z}_s = \mathcal{E}_\phi(\mathbf{M}) \in \mathbb{R}^{C\times H\times W\times D}, \qquad \mathbf{Z}_1 = \mathrm{stack}_{r=1}^{T}\,\mathcal{E}_\phi(\mathbf{Y}_r) \in \mathbb{R}^{C\times T\times H\times W\times D}. \tag{2}$$

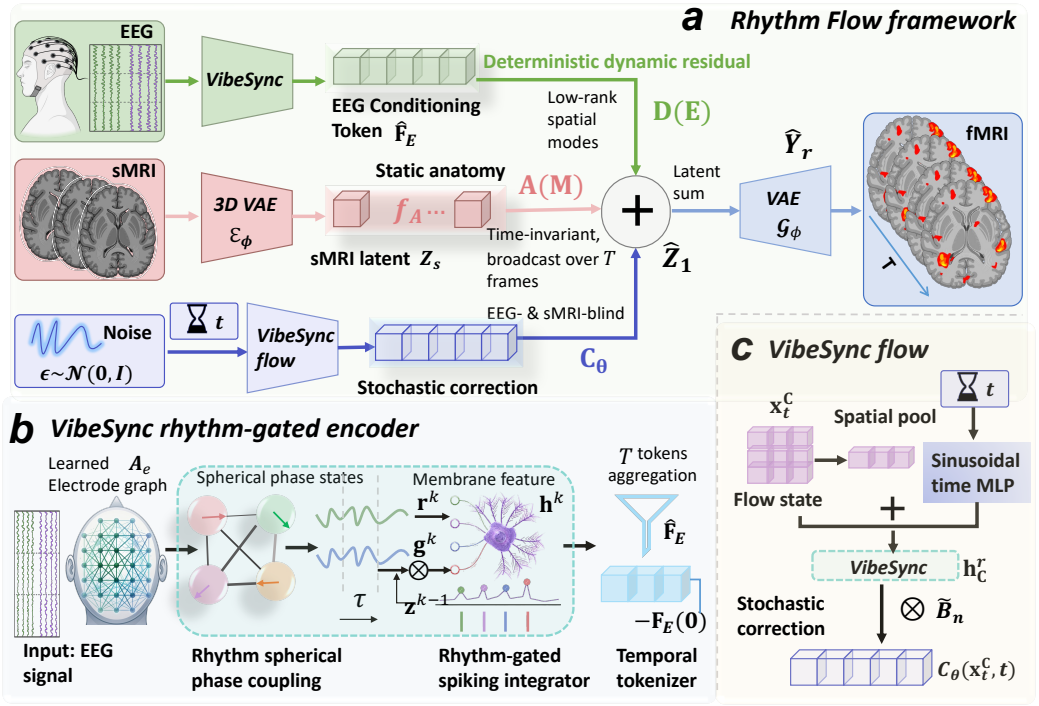

Figure 1: Overview of *Rhythm Flow*. **(a)** The 4-D fMRI latent includes three additive terms: a time-invariant anatomy branch $\mathbf{A}$ driven by sMRI, an EEG-driven deterministic residual $\mathbf{D}$, and a stochastic correction flow $\mathbf{C}_\theta$. **(b)** The *VibeSync* encoder couples spherical phase states over a learned electrode graph with a rhythm-gated spiking integrator, and aggregates the membrane features into T zero-referenced EEG tokens. **(c)** $\mathbf{C}_\theta$ uses the *VibeSync* backbone, driven by the pooled flow state and flow time t.

Let $\mathrm{P}_0(\mathbf{X}) = \mathbf{X} - \mathrm{mean}_T(\mathbf{X})$ denote the projection onto the zero-temporal-mean subspace. We decompose the target into a static and a dynamic part,

$$\mathbf{A}^\star = \mathrm{mean}_T(\mathbf{Z}_1), \qquad \mathbf{R}^\star = \mathrm{P}_0(\mathbf{Z}_1), \tag{3}$$

so that $\mathbf{Z}_1 = \mathbf{A}^\star + \mathbf{R}^\star$, with $\mathbf{A}^\star$ broadcast along time. Since $\mathbf{R}^\star$ has zero temporal mean, it is orthogonal to every time-invariant tensor, $\langle \mathbf{A}^\star, \mathbf{R}^\star \rangle = 0$ (Appendix C.2). This removes gradient competition between the anatomy branch, which predicts $\mathbf{A}^\star$, and the residual terms $\mathbf{D}$ and $\mathbf{C}_\theta$, which jointly predict $\mathbf{R}^\star$.

Static anatomy branch. The temporal mean of fMRI mainly reflects subject-specific anatomy and tissue contrast, so a 3-D residual convolutional network f_A predicts $\mathbf{A} = f_A(\mathbf{Z}_s)$ from the sMRI latent alone. Since f_A receives neither EEG nor temporal information, $\mathbf{A}$ is time-invariant by construction and is broadcast over the T frames.

2.2 EEG PATHWAY: VIBESYNC ENCODER AND EEG RESIDUAL

VibeSync treats the N electrodes as interacting units that each couple an oscillatory phase state with a spiking state. It consists of (1) rhythm spherical phase coupling (RSPC) over a learned electrode graph (Dan & Wu, 2026), (2) a rhythm-gated spiking integrator (Yeung & Strogatz, 1999; Yan et al., 2025), and (3) a temporal tokenizer with zero-referencing (details in Appendix C.3).

Rhythm spherical phase coupling. The classic Kuramoto coupling drives pairwise phase differences toward synchrony, which is too strong an inductive bias for EEG (Section 3.3). Following the vector oscillators of Miyato et al. (2025), RSPC places the phase state $\boldsymbol{\theta}_i^k$ of each electrode i on the

unit sphere and updates it as

$$\boldsymbol{\theta}_i^k = \mathrm{Normalize}\Big(\boldsymbol{\theta}_i^{k-1} + \Delta t\Big[\boldsymbol{\Omega}\,\boldsymbol{\theta}_i^{k-1} + \Pi_{\boldsymbol{\theta}_i^{k-1}}\Big(\textstyle\sum_j A_{e,ij}\,\boldsymbol{\theta}_j^{k-1} + \gamma_i^k \mathbf{1}\Big)\Big]\Big), \tag{4}$$

where γ_i^k is the EEG drive at step k, $\mathbf{A}_e = [\mathbf{P} + \mathbf{P}^\top]_+$ is a learnable symmetric electrode affinity, $\boldsymbol{\Omega}$ is a learnable rotation that sets the intrinsic frequencies, and $\Pi_{\mathbf{x}}(\mathbf{y}) = \mathbf{y} - (\mathbf{x}^\top\mathbf{y})\mathbf{x}$ is the tangent-space projection. Electrodes thus interact only through tangential updates, without a prescribed phase relationship.

Rhythm-gated spiking integrator. The current phase provides the spiking input, and the phase delayed by τ steps gates the membrane update,

$$\begin{aligned} &q_i^k = \tfrac{1}{2}\big[1 + \sin\bar{\theta}_i^{k-\tau}\big], \quad r_i^k = q_i^k\, s_i^k, \quad \mathbf{g}^k = \sigma(\mathbf{W}_g\mathbf{q}^k + \mathbf{b}_g), \\ &\mathbf{u}^k = (1 - \mathbf{g}^k) \odot \mathbf{u}^{k-1} + \mathbf{g}^k \odot \mathrm{LIF}\left(\mathbf{u}^{k-1}, \mathbf{r}^k, \mathbf{z}^{k-1}\right), \quad \mathbf{z}^k = \mathbb{I}\big[\mathbf{u}^k > v_{\mathrm{th}}\big], \end{aligned} \tag{5}$$

where $\bar{\theta}_i^k$ and s_i^k are the averages of $\boldsymbol{\theta}_i^k$ and $\sin\boldsymbol{\theta}_i^k$ over phase dimensions, bold symbols stack per-electrode quantities, and $\mathrm{LIF}(\cdot)$ is a leaky integrate-and-fire update with multi-branch dendritic input. The rhythm gate thus determines when each membrane integrates new input, which ties spiking to oscillatory timing. Gradients pass through the spike threshold via a surrogate derivative (Neftci et al., 2019).

Temporal tokenizer with zero-referencing. A spiking readout layer maps $\mathbf{z}^k$ to membrane features, which temporal convolution and adaptive average pooling aggregate into T tokens $\mathbf{F}_E \in \mathbb{R}^{T\times d_e}$. The tokens are not constrained to align with individual TRs. Since $\mathbf{D}$ must vanish when EEG carries no information, we zero-reference the tokens against the encoder response to an all-zero recording, $\widehat{\mathbf{F}}_E = \mathbf{F}_E(\mathbf{E}) - \mathbf{F}_E(\mathbf{0})$, so that $\widehat{\mathbf{F}}_E \equiv \mathbf{0}$ whenever $\mathbf{E} \equiv \mathbf{0}$.

EEG residual. $\mathbf{D}$ maps $\widehat{\mathbf{F}}_E$ to a whole-brain latent trajectory through a low-rank spatiotemporal expansion. A temporal mixer reads off K mode coefficients per token, $\boldsymbol{\alpha} = \mathbf{W}_\alpha f_{\mathrm{mix}}(\widehat{\mathbf{F}}_E) \in \mathbb{R}^{T\times K}$, which we treat as temporal derivatives and integrate over the token axis,

$$\mathbf{c} = \mathrm{P}_0\left(\mathrm{cumsum}_T\,\mathrm{P}_0(\boldsymbol{\alpha})\right)\big|_{\mathbf{E}} - \mathrm{P}_0\left(\mathrm{cumsum}_T\,\mathrm{P}_0(\boldsymbol{\alpha})\right)\big|_{\mathbf{0}}, \quad \mathbf{D} = \mathrm{P}_0\left(\tfrac{\kappa}{\sqrt{K}}\textstyle\sum_{k=1}^K \mathbf{c}_{:,k} \otimes \widetilde{\mathbf{S}}_k\right), \tag{6}$$

where the second term of $\mathbf{c}$ re-applies zero-referencing at the coefficient level, $\widetilde{\mathbf{S}}_k$ are the RMS-normalized versions of learnable whole-brain spatial modes $\mathbf{S}_k$ shared across subjects, and κ is a fixed scale. EEG thus parameterizes how the latent changes from one step to the next, so rhythmic input can accumulate into slow drifts, in line with the hypothesis that oscillatory neural activity shapes hemodynamics on a slower timescale (Jacob et al., 2025; Fultz et al., 2019).

2.3 VibeSync correction flow

Since $\mathbf{D}$ is deterministic and low-rank, it cannot represent all of $\mathbf{R}^\star$. A stochastic correction flow $\mathbf{C}_\theta$ models the remainder and depends only on the flow state and the flow time t. This modality blindness is essential, since a flow that observes EEG could absorb $\mathbf{D}$ and reintroduce the identifiability problem.

VibeSync backbone. Rhythmic spiking integration is also a plausible inductive bias for population-level BOLD dynamics, so $\mathbf{C}_\theta$ reuses the *VibeSync* architecture (Equations (4)–(5)) and replaces the EEG drive with a projection of the spatially pooled flow state plus a flow-time embedding over N_C latent nodes. A learned whole-brain basis maps the resulting membrane trajectories $\mathbf{h}_{C,n}^{1:T}$ back to the latent space, $\mathbf{C}_\theta(\mathbf{x}_t^C, t) = \sum_n \mathbf{h}_{C,n}^{1:T} \otimes \widetilde{\mathbf{B}}_n$.

Correction coordinate with stop-gradient. $\mathbf{C}_\theta$ is trained on a correction coordinate that removes $\mathbf{D}$ from the interpolated state,

$$\mathbf{x}_t^C = \mathrm{P}_0\left(\mathbf{x}_t - t\ \mathrm{sg}(\mathbf{D})\right), \qquad \mathbf{x}_t = (1-t)\boldsymbol{\epsilon} + t\,\mathbf{R}^\star, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{7}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^{C\times T\times H\times W\times D}$ is Gaussian noise with the shape of $\mathbf{R}^\star$ and $\mathrm{sg}(\cdot)$ is the stop-gradient operator. Since $\mathbf{x}_0^C = \mathrm{P}_0(\boldsymbol{\epsilon})$ and $\mathbf{x}_1^C = \mathbf{R}^\star - \mathrm{sg}(\mathbf{D})$, this coordinate defines a straight path from noise to the correction target $\mathbf{C}^\star = \mathbf{R}^\star - \mathrm{sg}(\mathbf{D})$. Detaching $\mathbf{D}$ is the central safeguard of the design. Without it, $\mathbf{C}_\theta$ could learn to cancel whatever $\mathbf{D}$ produces and absorb its role while leaving $\mathbf{D} + \mathbf{C}_\theta$ unchanged. With $\mathbf{D}$ detached, $\mathbf{D}$ is supervised only by its own objectives (Section 2.4).

2.4 TRAINING OBJECTIVE AND INFERENCE

Composition and supervision. The final prediction is composed and decoded frame-wise as

$$\widehat{\mathbf{Z}}_1 = \mathbf{A} \otimes \mathbf{1}_T + \mathrm{P}_0\left(\mathbf{D} + \mathbf{C}_\theta\right), \qquad \widehat{\mathbf{Y}}_r = \mathcal{G}_\phi\big(\widehat{\mathbf{Z}}_{1,r}\big), \tag{8}$$

where P_0 enforces the zero temporal mean of Equation (1). We supervise $\mathbf{A}$ and $\mathbf{D}$ directly with $\mathcal{L}_{\text{static}} = \mathrm{MSE}(\mathbf{A}, \mathbf{A}^\star)$ and $\mathcal{L}_{\mathrm{D}} = \mathrm{NMSE}(\mathbf{D}, \mathbf{R}^\star) + \lambda_\Delta \mathrm{NMSE}(\Delta_T \mathbf{D}, \Delta_T \mathbf{R}^\star)$, where NMSE is the normalized MSE and Δ_T is the adjacent-frame difference. $\mathbf{C}_\theta$ predicts the clean endpoint $\mathbf{C}^\star$ and is trained by flow matching,

$$\mathcal{L}_{\text{flow}} = \mathbb{E}_{t,\boldsymbol{\epsilon}}\Big\| \frac{\mathbf{C}_\theta(\mathbf{x}_t^C, t) - \mathbf{x}_t^C}{1-t} - \big(\mathbf{C}^\star - \mathrm{P}_0(\boldsymbol{\epsilon})\big) \Big\|_2^2, \tag{9}$$

which is disabled for the first three epochs so that $\mathbf{A}$ and $\mathbf{D}$ fit the target first.

EEG-conditioning evaluation. For each sample, we construct two controls with the same anatomy: (1) a *within-run* control, i.e., the EEG of a non-overlapping window of the same run, and (2) a *cross-subject* control, i.e., a task-, run-, and window-matched EEG segment of another subject. The sampler of $\mathbf{C}_\theta$ starts from $\mathrm{P}_0(\boldsymbol{\epsilon})$ and never receives $\mathbf{D}$ (Equation (17)), so under shared noise, $\mathbf{C}_\theta$ yields identical outputs for all EEG conditions, and

$$\widehat{\mathbf{Z}}_1^{\text{correct}} - \widehat{\mathbf{Z}}_1^{\text{control}} = \mathbf{D}(\mathbf{E}^{\text{correct}}) - \mathbf{D}(\mathbf{E}^{\text{control}}). \tag{10}$$

Let $\rho(\mathbf{X})$ denote the cosine similarity between $\Delta_T \mathbf{X}$ and $\Delta_T \mathbf{R}^\star$, averaged over frames. We score each EEG condition by ρ of the composed prediction, denoted ρ^+ for the matched EEG and $\rho^- \in \{\rho^-_{\text{run}}, \rho^-_{\text{cross}}\}$ for the two controls, and select checkpoints by the *EEG-conditioning gap*, i.e., the validation mean of ρ^+ minus the mean control score, computed on the EEG residual $\mathbf{D}$ (on HBN, which has one EEG segment per subject, the cross-subject control only). Applied to $\mathbf{D}$, the same score also provides a training signal against an EEG-insensitive $\mathbf{D}$,

$$\mathcal{L}_{\text{EEG}} = \mathbb{E}\Big[1 - \rho\big(\mathbf{D}(\mathbf{E})\big) + \lambda_r\, \beta \log\Big(1 + \exp\Big(\frac{\rho\big(\mathbf{D}(\mathbf{E}^-)\big) - \rho\big(\mathbf{D}(\mathbf{E})\big) + m}{\beta}\Big)\Big)\Big], \tag{11}$$

where $\mathbf{E}^-$ is a control EEG, and the second term penalizes any control whose score falls within a margin m of the matched EEG.

Total objective and inference. The total objective is

$$\mathcal{L} = \mathcal{L}_{\text{static}} + \mathcal{L}_{\mathrm{D}} + \mathcal{L}_{\text{flow}} + \lambda_E \mathcal{L}_{\text{EEG}} + \lambda_R \mathcal{L}_{\text{reg}}, \tag{12}$$

where $\mathcal{L}_{\text{reg}}$ combines orthogonality and temporal-smoothness regularizers on $\mathbf{D}$ (Appendix C.5). At inference, $\mathbf{A}$ and $\mathbf{D}$ are computed in a single forward pass, and only $\mathbf{C}_\theta$ is sampled, by Euler integration from $\mathrm{P}_0(\boldsymbol{\epsilon})$, before the prediction is composed via Equation (8). Sampling variance under a fixed EEG therefore arises only from $\boldsymbol{\epsilon}$.

3 EXPERIMENTS

3.1 EXPERIMENTAL SETUP

Data. We include two types of data in our work. (1) *Simultaneous cohorts.* We use ds002336 and ds002338 (Lioi et al., 2020), with 6 and 11 training subjects and 2 and 3 validation subjects, respectively. The EEG is artifact-corrected and aligned to the R128 volume triggers with a hemodynamic lag of 3 TR. ds002336 uses a 32-second EEG window at TR $= 2.0$ s, and ds002338 a 16-second window at TR $= 1.0$ s. Joint training (`multi`) uses both cohorts (17 subjects, 1634 windows) and crops the EEG of ds002336 to 16 seconds to share one encoder input. (2) *HBN.* We use the tri-modal HBN cohort (Alexander et al., 2017), which pairs resting-state EEG from the OpenNeuro HBN releases with T1-weighted sMRI and resting-state fMRI. The 1152 subjects with all three modalities are split into 1021 training, 39 validation, and 92 development subjects without subject overlap (Appendix A). We evaluate temporal generalization (training subjects, a later 12.8-second window), subject generalization (unseen subjects, the first window), and joint generalization (unseen subjects, a later window). EEG and fMRI in HBN are paired at the subject level but not time-locked, and

one 30-second EEG segment conditions both windows of a subject. All volumes are resampled to $80 \times 80 \times 48$ in subject space.

Metrics. We report MSE and PSNR for voxel fidelity and RSA-ROI for representational structure, both computed on decoded volumes. RSA-ROI correlates the region-by-region cosine-similarity matrices of the prediction and the target over 64 fixed regions of the voxel grid. For EEG specificity, we report the paired advantage $\rho^+ - \bar{\rho}^-$ of matched EEG over the controls (Section 2.4).

Implementation. On the simultaneous cohorts, each model is trained for up to 100 epochs with early stopping (patience 8) under three seeds and with auxiliary loss of weight 0.1 (connectivity preservation; Appendix C.5). On HBN, all models use the same data split: they are trained on the 1021 training subjects, evaluated on these subjects for temporal generalization, and evaluated on the same 92 development subjects for subject and joint generalization. *Rhythm Flow* is trained from scratch for up to 300 epochs, and we keep the checkpoint with the largest EEG-conditioning gap on the 39 validation subjects. The single-flow generators are trained for up to 5000 epochs, and the mean template averages the fMRI latents of all training subjects. Optimization details are given in Appendix C.6.

Published methods. Table 3 (Appendix B) summarizes five published EEG-to-fMRI methods: (1) Sparse CCA (Deligianni et al., 2014), which predicts fMRI connectomes from source-space EEG connectomes, (2) the attentional graph model (Calhas & Henriques, 2023), which synthesizes single whole-brain fMRI volumes, (3) NeuroBOLT (Li et al., 2024), which predicts regional BOLD time series, (4) UnEBOLT (Li et al., 2026), which jointly reconstructs regional time series and functional connectivity, and (5) Diffusion–Mamba (Ma et al., 2025), which generates fMRI from EEG and structural images. Since these studies differ in cohorts, targets, preprocessing, and metrics, their values serve as reference only.

3.2 Results on simultaneous EEG–fMRI

Table 1 reports RSA-ROI and the paired advantage of matched EEG, and Figure 2 shows the per-subject permutation tests (Appendix D).

Table 1: Results on the simultaneous cohorts over three seeds. `full` is *Rhythm Flow* trained on one cohort, `multi` is trained on both cohorts (6 + 11 subjects) and evaluated on each, and `no_eeg` removes the EEG pathway ($\mathbf{A}+\mathbf{C}_\theta$). RSA-ROI is the mean over seeds with the range in parentheses. The advantage $\rho^+ - \bar{\rho}^-$ of matched EEG over the within-run null (all other windows of the same run) and the cross-subject null (up to 32 same-task windows of other subjects) is averaged over seeds and validation subjects. Brackets give the number of subject–seed tests with $p < 0.05$ (5000 permutations). Since each subject recurs across seeds, these tests are not independent.

				Advantage ($\times 10^{-3}$)	
Model	**Cohort**	$N_{\mathbf{train}}$	**RSA-ROI↑**	**within-run**	**cross-subject**
`full`	ds002336	6	0.090 (0.018–0.143)	+1.7 [6/6]	+1.8 [6/6]
`full`	ds002338	11	0.284 (0.256–0.318)	+1.2 [4/9]	+1.1 [3/9]
`multi`	ds002336	17	0.123 (0.082–0.159)	+0.2 [1/6]	+0.2 [2/6]
`multi`	ds002338	17	0.285 (0.259–0.322)	+1.0 [4/9]	+0.9 [3/9]
`no_eeg`	ds002336	6	0.082 (0.010–0.157)	–	–
`no_eeg`	ds002338	11	0.234 (0.184–0.281)	–	–

EEG specificity. For `full` on ds002336, matched EEG outperforms both nulls with $p < 0.05$ for both validation subjects in every seed. On ds002338, its advantage is positive for every subject and seed, and xp216 separates from both nulls with $p \leq 0.002$ in every seed (Figure 2a). We attribute this subject dependence to the spatial modes of $\mathbf{D}$, which are shared across subjects (Equation (6)) and thus capture subjects whose EEG–BOLD coupling projects onto this basis. Since the controls share $\mathbf{A}$ and $\mathbf{C}_\theta$, RSA-ROI changes little across EEG conditions, so we assess specificity with the paired statistic, which isolates $\mathbf{D}$ by construction (Equation (10)).

Joint training. Joint training raises the RSA-ROI of ds002336 from 0.090 to 0.123 and keeps that of ds002338 at the level of `full` (0.285 against 0.284), and xp216 remains significant against both

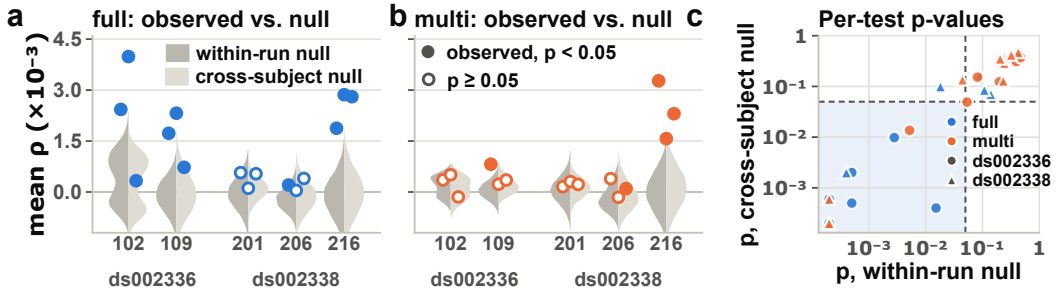

Figure 2: Permutation test for EEG specificity on the simultaneous cohorts for **(a)** `full` and **(b)** `multi`. For each validation subject, the observed mean ρ of matched EEG in each seed (dots, filled when $p < 0.05$ against the within-run null) is shown against the within-run null (left, darker) and the cross-subject null (right, lighter), each pooled over three seeds of 5000 permutations. **(c)** The two one-sided p-values of every subject–seed test on logarithmic axes. Dashed lines mark $p = 0.05$, and the shaded square marks significance under both nulls.

nulls in every seed (Figure 2b). On ds002336, the within-run advantage drops from $+1.7 \times 10^{-3}$ to $+0.2 \times 10^{-3}$. We attribute this drop to the shared input format. With $T = 16$ tokens for both cohorts, one token indexes 1 s of BOLD in ds002338 and 2 s in ds002336, so the cumulative integration and the spatial modes of $\mathbf{D}$ must serve two timescales. In addition, the cropped EEG of ds002336 covers only the first half of its 32-second fMRI window.

Out-of-fold cross-validation. Since two validation subjects cannot support a population claim, we repeat the ds002336 experiment under five-fold subject cross-validation with two sealed test subjects per fold (Appendix E). Over the ten test subjects, the within-run advantage of `full` is positive for eight subjects. Without the correction flow (`no_flow`, i.e., $\mathbf{A} + \mathbf{D}$), it is positive for nine subjects ($+0.78 \times 10^{-3}$, $p = 0.024$; cross-subject null, $p = 0.032$). The correction flow in turn raises the out-of-fold RSA-ROI from 0.023 to 0.135 at the selected checkpoints. The two terms thus fulfill the roles assigned by the factorization: $\mathbf{D}$ carries the EEG specificity, and $\mathbf{C}_\theta$ supplies the residual dynamics that $\mathbf{D}$ does not explain.

3.3 Results on HBN

Having established EEG specificity on simultaneous cohorts (Section 3.2), we now shift our focus to evaluating generation quality using the HBN dataset, where EEG and fMRI were acquired asynchronously. Table 2 compares *Rhythm Flow* with six single-flow generators, a mean template, and a region-level predictor (same as the output format of NeuroBOLT). The single-flow generators follow the design that motivated our factorization, i.e., one flow-matching generator conditioned on both modalities, with *VibeSync* tokens injected through cross-attention and the sMRI latent through adaptive normalization. They share all hyperparameters and are evaluated at a matched budget of 3600 to 3900 epochs.

Main comparison. *Rhythm Flow* consistently outperforms all baseline and ablated models across all evaluation metrics in Table 2. On subject generalization, it achieves the highest RSA-ROI of 0.469 and the lowest voxel-level error (MSE 0.0104, PSNR 37.5 dB). Notably, since RSA-ROI does not remove regional means, the mean template already provides a formidable baseline for subject generalization (0.413), surpassing several single-flow variants. *Rhythm Flow* exceeds this baseline on 33 of the 53 subjects with a defined score (Figure 3c). This gain highlights the core advantage of our factorization design: single-flow architectures struggle with an optimization bottleneck, as they must continuously reconstruct the highly individualized, static spatial topography from noise at every time step. In contrast, our anatomy branch explicitly regresses the temporal mean $\mathbf{A}^\star$, effectively anchoring the subject's spatial backbone and freeing the generative capacity to model dynamic fluctuations.

Furthermore, *Rhythm Flow* outperforms `RSPC`, which leads the single-flow variants on temporal and joint generalization (0.160 and 0.114), reaching RSA-ROIs of 0.173 and 0.160, respectively. This indicates that extracting the low-rank EEG residual, prevents the deterministic, task-specific

EEG representations from being washed out by the inherently high-variance fMRI noise during the generative process. Figure 3 illustrates both advantages on an unseen development subject. The voxel-wise error maps (Figure 3a) show that *Rhythm Flow* yields lower absolute errors than baseline (i.e., `kuramoto` model in Table 2) across the axial, coronal, and sagittal slices. Meanwhile, the region similarity matrices (Figure 3b) show that *Rhythm Flow* reproduces the region similarity structure of the target (RSA-ROI 0.94 against 0.70 for `kuramoto`). Both observations are consistent with the factorization, where the anatomy branch recovers the static spatial pattern that a single flow need regenerate from noise. It is worth noting that the 0.469 average RSA-ROI for subject generalization in Table 2 is a penalized metric over all 92 development subjects. For 39 of the 92 development subjects, the target BOLD fluctuates only about half as much as for the other subjects, and the prediction of *Rhythm Flow* varies even less over time. Their region similarities are then nearly uniform, the correlation is undefined, and the benchmark sets RSA-ROI to zero. For the remaining 53 subjects (Figure 3c), *Rhythm Flow* yields an average RSA-ROI of 0.73, remaining competitive with the mean template (0.69) on this subset.

Table 2: Comparison and ablation on HBN. RSA-ROI is given for temporal (temp), subject (subj), and joint generalization, and MSE and PSNR for subject generalization. The single-flow generators condition one flow on both modalities. `kuramoto` uses sine-difference Kuramoto coupling, and the other variants modify it: `smri_only` and `eeg_only` zero one modality, `no_rhythm` removes the rhythm gate, `no_kuramoto` removes the coupling, and `RSPC` replaces it with the spherical coupling of Equation (4). The mean template returns the mean training fMRI latent for every subject. The ROI predictor maps EEG to 64 regional BOLD time courses with the *VibeSync* encoder and a per-frame decoder in the output format of NeuroBOLT, so its MSE and PSNR are not comparable. *Rhythm Flow* is reported as mean ± standard deviation over three seeds. Best results are in bold.

	RSA-ROI↑			Fidelity (subj.)	
Model	**temp**	**subj**	**joint**	**MSE↓**	**PSNR↑**
`smri_only`	0.125	0.264	0.069	0.0128	36.7
`eeg_only`	0.151	0.377	0.124	0.0188	35.1
`kuramoto`	0.086	0.229	0.042	0.0126	36.6
`no_rhythm`	0.069	0.078	0.037	0.0124	36.8
`no_kuramoto`	0.093	0.458	0.106	0.0122	36.8
`RSPC`	0.160	0.421	0.114	0.0123	36.8
Mean template	0.061	0.413	0.076	0.0109	37.4
ROI predictor	0.098	0.319	0.108	–	–
NeuroBOLT	0.069 ± 0.008	0.407 ± 0.037	0.072 ± 0.007	–	–
Rhythm Flow	$\mathbf{0.173 \pm 0.015}$	$\mathbf{0.469 \pm 0.036}$	$\mathbf{0.160 \pm 0.011}$	**0.0104**	**37.5**

Ablation study. The single-flow variants yield three findings. First, the two modalities compete when they share one generative flow. `eeg_only` outperforms `smri_only` in all three settings, yet `kuramoto`, which receives both modalities, falls below both single-modality variants. Since a single flow places both modalities on one undivided target, the gradient of one modality can suppress the other. Our orthogonal decomposition (Equation (3)) removes this competition by assigning each modality its own target subspace. Second, the coupling law determines performance. Both `no_kuramoto` and `RSPC`, which couples electrodes only through tangential updates, surpass `kuramoto` in all three settings, indicating that sine-difference synchronization imposes an overly inductive bias on EEG. Third, the rhythm gate is essential. Without it, `no_rhythm` falls below every other single-flow variant in all three settings. Since the rhythm gate determines when each membrane integrates new input (Equation (5)), removing it detaches the spiking pathway from oscillatory timing, so the tokens lose the phase structure that conditions the generator.

Taken together, these findings motivate the design of *Rhythm Flow*, which assigns the two modalities separate targets, replaces sine-difference synchronization with spherical phase coupling, and gates spiking integration by rhythm.

Discussion and Limitations. While *Rhythm Flow* successfully factorizes the multimodal latent space and isolates EEG-specific temporal dynamics, our study presents certain limitations. First, the

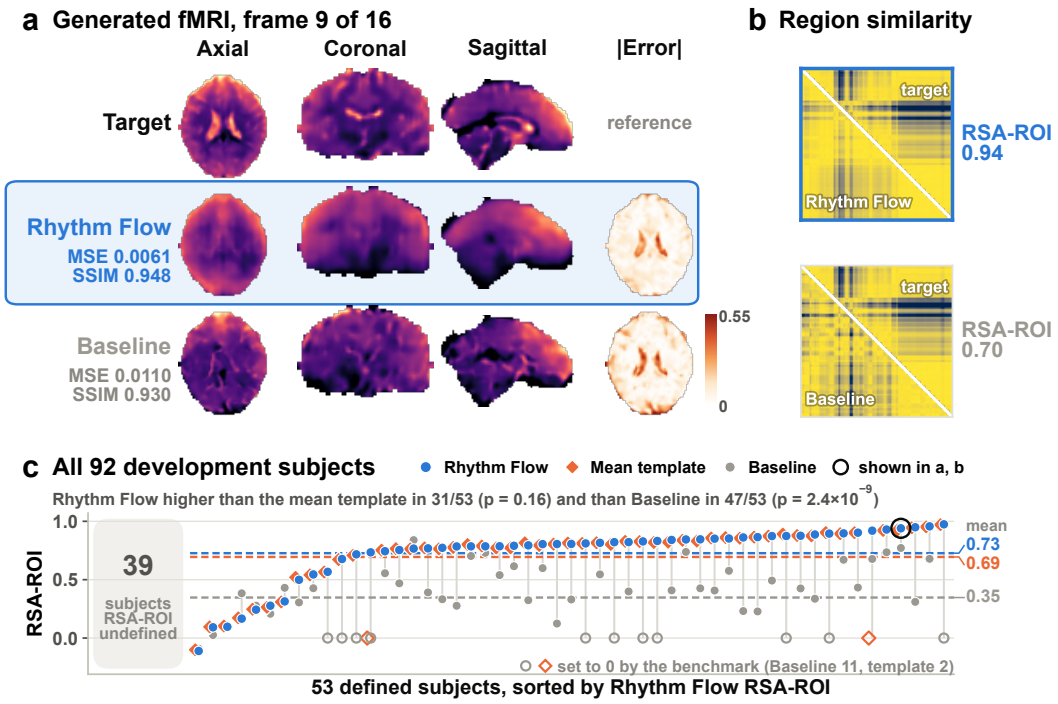

Figure 3: Generated fMRI for an unseen HBN subject (subject generalization). **(a)** Axial, coronal, and sagittal slices of frame 9 for the target, *Rhythm Flow*, and the `kuramoto` single-flow generator (Baseline), with voxel-wise absolute error on a shared color scale and the window-level MSE and SSIM (blue marks the better model). **(b)** Region-by-region cosine similarity over the 64 RSA-ROI regions, ordered by hierarchical clustering of the target. The upper triangle shows the target and the lower triangle the model, so symmetry about the diagonal indicates a faithful reproduction. **(c)** Subject-generalization RSA-ROI of all 92 development subjects for *Rhythm Flow* in *Seed #1*, the mean template, and Baseline. For 39 subjects (shaded), the predicted region similarity of *Rhythm Flow* is too uniform for a correlation, so its RSA-ROI is undefined. On the remaining 53 subjects, *Rhythm Flow* has the higher mean (0.73, against 0.69 for the template and 0.35 for Baseline; Wilcoxon $p = 0.16$ versus the template) and exceeds Baseline on 47 subjects.

simultaneous EEG–fMRI cohorts provide only two or three validation subjects each. Although we partly compensate for this through rigorous subject cross-validation, evaluating on larger simultaneous datasets is necessary to further solidify the population-level claims of EEG specificity. Second, our representational evaluation (RSA-ROI) relies on a fixed voxel partition, which does not measure fine-grained atlas-level recovery. Finally, because EEG specificity can only be strictly verified on time-locked recordings, the larger HBN dataset could only be used to evaluate general generation quality. In future work, we plan to adopt atlas-based representational metrics and introduce subject-adaptive spatial modes into $\mathbf{D}$ to better capture subject-specific EEG–BOLD coupling mechanisms.

4 Conclusion

In this work, we present *Rhythm Flow*, which factorizes the fMRI latent into a time-invariant anatomy branch, a deterministic EEG-driven residual, and a modality-blind stochastic correction flow, and we devise *VibeSync*, which encodes EEG through spherical phase coupling and rhythm-gated spiking integration. Since the EEG-driven residual is an explicit function of EEG alone, its specificity can be tested on any trained model. On two simultaneous cohorts, matched EEG outperforms the within-run and cross-subject controls, and under subject cross-validation, the EEG-driven residual without the correction flow is positive for nine of ten sealed test subjects. On HBN, *Rhythm Flow* outperforms the single-flow generators, the ROI predictor, and the mean template in both representational similarity and voxel fidelity.

AI USE STATEMENT

In this work, we used generative AI tools for coding and writing polishing. We have not used generative AI tools for figure generation, dataset construction, or ideation. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by the first author. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses only public, de-identified datasets (OpenNeuro ds002336 and ds002338, and the Healthy Brain Network), which the original studies collected with informed consent (Lioi et al., 2020; Alexander et al., 2017) and which we use under their data-use terms. We make no attempt to re-identify participants.

REPRODUCIBILITY STATEMENT

Section 2 and Appendix C specify the architecture, losses, and sampler, Table 4 lists all hyperparameters, and Section 3.1 and Appendix A describe the data and splits. Code will be published after acceptance.

REFERENCES

Lindsay M Alexander, Jasmine Escalera, Lei Ai, Charissa Andreotti, Karina Febre, Alexander Mangone, Natan Vega-Potler, Nicolas Langer, Alexis Alexander, Meagan Kovacs, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific data*, 4(1):170181, 2017.

David Calhas and Rui Henriques. EEG to fMRI synthesis benefits from attentional graphs of electrode relationships. In *Proceedings of the 8th Machine Learning for Healthcare Conference*, volume 219 of *Proceedings of Machine Learning Research*, pp. 76–93. PMLR, 2023. URL https://proceedings.mlr.press/v219/calhas23a.html.

Tingting Dan and Guorong Wu. From cortical synchronous rhythm to brain inspired learning mechanism: An oscillatory spiking neural network with time-delayed coordination. *arXiv preprint arXiv:2605.01656*, 2026.

Fani Deligianni, Maria Centeno, David W. Carmichael, and Jonathan D. Clayden. Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. *Frontiers in Neuroscience*, 8:258, 2014. doi: 10.3389/fnins.2014.00258.

Nina E Fultz, Giorgio Bonmassar, Kawin Setsompop, Robert A Stickgold, Bruce R Rosen, Jonathan R Polimeni, and Laura D Lewis. Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep. *Science*, 366(6465):628–631, 2019.

Gary H. Glover. Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America*, 22(2):133–139, 2011.

Evan M. Gordon et al. Precision functional mapping of individual human brains. *Neuron*, 95(4): 791–807, 2017.

Leandro PL Jacob, Sydney M Bailes, Stephanie D Williams, Carsen Stringer, and Laura D Lewis. Brainwide hemodynamics predict eeg neural rhythms across sleep and wakefulness in humans. *PLOS Computational Biology*, 21(9):e1013497, 2025.

Yamin Li, Ange Lou, Ziyuan Xu, Shengchao Zhang, Shiyu Wang, Dario J. Englot, Soheil Kolouri, Daniel Moyer, Roza G. Bayrak, and Catie Chang. NeuroBOLT: Resting-state EEG-to-fMRI synthesis with multi-dimensional feature mapping. In *Advances in Neural Information Processing Systems*, volume 37, pp. 23378–23405, 2024. doi: 10.52202/079017-0736.

Yamin Li, Ange Lou, Chang Li, Shiyu Wang, Haatef Pourmotabbed, Ziyuan Xu, Shengchao Zhang, Dario J. Englot, Soheil Kolouri, Daniel Moyer, Roza G. Bayrak, and Catie Chang. UnEBOLT: A unified model for EEG-to-BOLD translation and functional connectivity reconstruction. In *Proceedings of the 9th International Conference on Medical Imaging with Deep Learning*, volume 315 of *Proceedings of Machine Learning Research*, pp. 2338–2351. PMLR, 2026. URL https://proceedings.mlr.press/v315/li26g.html.

Giulia Lioi, Claire Cury, Lorraine Perronnet, Marsel Mano, Elise Bannier, Anatole Lécuyer, and Christian Barillot. Simultaneous EEG-fMRI during a neurofeedback task, a brain imaging dataset for multimodal data integration. *Scientific Data*, 7:173, 2020. doi: 10.1038/s41597-020-0498-3.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Zhenren Ma, Jialin Li, Jiabao Wen, and Wenjing Yan. A deep generative framework based on a diffusion model integrated with Mamba block in modality transferring from resting-state EEG to fMRI. In *Proceedings of the 2025 14th International Conference on Bioinformatics and Biomedical Science*, pp. 130–134. ACM, 2025. doi: 10.1145/3778067.3778084.

Christoph M. Michel and Denis Brunet. EEG source imaging: A practical review of the analysis steps. *Frontiers in Neurology*, 10:325, 2019.

Takeru Miyato, Sindy Löwe, Andreas Geiger, and Max Welling. Artificial kuramoto oscillatory neurons. In *International Conference on Learning Representations*, volume 2025, pp. 44278–44322, 2025.

Sophia Mueller, Danhong Wang, Michael D. Fox, B. T. Thomas Yeo, Jorge Sepulcre, Mert R. Sabuncu, Rebecca Shafee, Jie Lu, and Hesheng Liu. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586–595, 2013.

Emre O. Neftci, Hesham Mostafa, and Friedemann Zenke. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63, 2019.

Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4): 229–239, 2001.

Yinsong Yan, Qu Yang, Yujie Wu, Hanwen Liu, Malu Zhang, Haizhou Li, Kay Chen Tan, and Jibin Wu. Efficient and robust temporal processing with neural oscillations modulated spiking neural networks. *Nature communications*, 16(1):8651, 2025.

M. K. Stephen Yeung and Steven H. Strogatz. Time delay in the Kuramoto model of coupled oscillators. *Physical Review Letters*, 82(3):648–651, 1999.

A COHORT PROVENANCE AND REPRESENTATIVE MODALITIES

Figure 4 profiles the 1152 HBN subjects used in our experiments, split into 1021 training, 39 validation, and 92 development subjects. They comprise every subject of the OpenNeuro HBN EEG releases with resting-state EEG, T1-weighted sMRI, and resting-state fMRI. Figure 5 illustrates the three modalities.

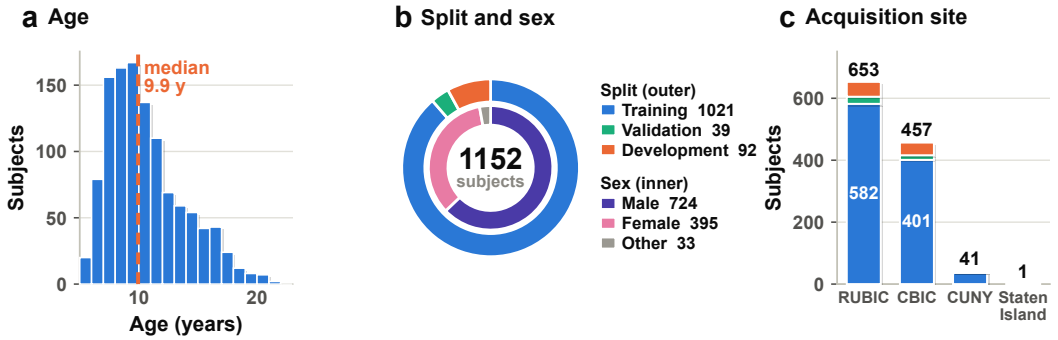

Figure 4: The 1152 HBN subjects used in our experiments. **(a)** Age distribution. **(b)** Split into training, validation, and development subjects (outer ring) and sex composition (inner ring). **(c)** Acquisition sites, stacked by split.

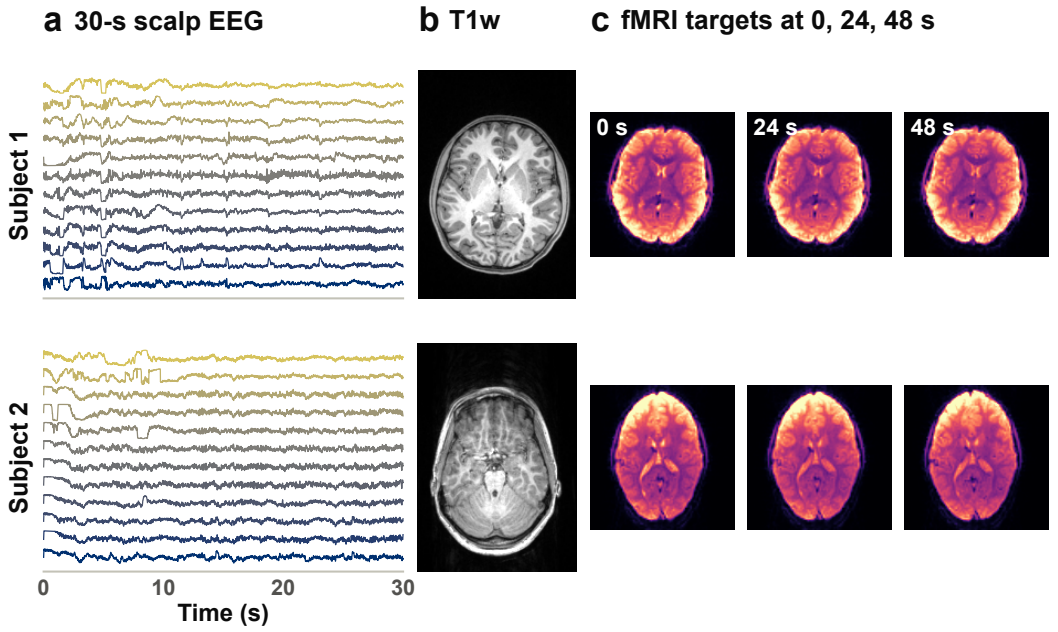

Figure 5: Representative modalities for two HBN training subjects, selected by an EEG-quality screen without reference to model outputs. **(a)** Twelve evenly spaced channels of the 30-second EEG. **(b)** Central axial T1-weighted sMRI slice. **(c)** Central axial fMRI slices at frames 0, 30, and 60 (0, 24, and 48 seconds at TR $= 0.8$ s). EEG and fMRI are acquired in separate sessions.

B PUBLISHED EEG-TO-FMRI METHODS

Table 3 lists the scope and the originally reported results of the published methods of Section 3.1.

Table 3: Scope and originally reported results of published EEG-to-fMRI methods, together with the setting of *Rhythm Flow*. Values remain tied to the datasets, splits, outputs, preprocessing, and metrics of each publication and do not form a controlled leaderboard. "Full vol." denotes voxel-wise whole-volume prediction and "Native anat." prediction in subject space. N reports subjects/scans when both are available. FC denotes functional connectivity, and NR denotes information that could not be verified from the publication.

Method	Dataset / task	N (subj./ scans)	Input	EEG win. (s)	Output	Full vol.	Native anat.	Original-paper result
Sparse CCA (Deligianni et al., 2014)	Deligianni, eyes-open rest	17 subjects	64-ch EEG, source-space FC	full scan	68/82-node fMRI connectome	No	N/A	For the WTS analysis, EEG→fMRI had lower geodesic error than the reverse direction in every band ($p < 10^{-5}$). Adding subcortical EEG connections improved cortical-fMRI prediction in δ ($p = 0.01$), with a trend in θ ($p = 0.08$).
Attentional graph (Calhas & Henriques, 2023)	NODDI / EOEC rest Oddball / visual detection CN-EPFL / decision making	10 10 20	64-ch EEG 43-ch EEG 64-ch EEG	20	single 3-D fMRI volume	Yes	No	Best proposed variant per dataset. NODDI: RMSE 0.3854 ± 0.0108, SSIM 0.4840 ± 0.0070. Oddball: RMSE 0.7014 ± 0.0855, SSIM 0.2004 ± 0.0172. CN-EPFL: RMSE 0.5166 ± 0.0560, SSIM 0.5222 ± 0.0877.
NeuroBOLT (Li et al., 2024)	in-house / eyes-closed rest in-house / auditory task	22/29 10/16	26-ch EEG (rest) 31-ch EEG (task)	16	regional and global BOLD traces	No	No	Resting-state intra-scan/held-out-subject mean Pearson $R = 0.531/0.473$ and average MSE $= 0.248/0.197$. Auditory held-out-subject average $R = 0.380/0.373/0.418/0.420$ for zero-shot, task-only, fine-tuned, and joint training, respectively.
UnEBOLT (Li et al., 2026)	same resting cohort	22/29	26-ch EEG	16	DiFuMo-64 traces + FC	No	No	Intra-scan/held-out-subject: Tcorr $0.416(0.166)/0.418(0.036)$, Pixcorr $0.490(0.176)/0.549(0.120)$, ConnMSE $0.201(0.148)/0.092(0.066)$, $\mathrm{F1}_{.25}$ $0.474(0.067)/0.545(0.074)$, and $\mathrm{F1}_{.50}$ $0.672(0.062)/0.690(0.047)$.
Diffusion–Mamba (Ma et al., 2025)	resting state, dataset details NR	NR	EEG + sMRI	NR	fMRI images + regional FC	NR	NR	SSIM $= 0.809$, PSNR $= 24.04$ dB, and Pearson $r = 0.53$ between FC matrices derived from generated and real fMRI.
Rhythm Flow	ds002336, ds002338 (simultaneous) HBN (Alexander et al., 2017) (separate)	6 / 11 train 1021 train	63-ch EEG + sMRI 128-ch EEG + sMRI	32 / 16 30	16-frame voxel-wise 4-D fMRI	Yes	Yes*	See Table 1 for the simultaneous cohorts and Table 2 for HBN.

Notes. NeuroBOLT recruited 24 resting-state participants and analyzed 22 subjects with 29 usable scans. Parentheses in the UnEBOLT row denote standard deviations. **Rhythm Flow* predicts in subject space but resamples each volume to $80 \times 80 \times 48$ and does not retain the native voxel lattice.

B.1 NeuroBOLT on HBN

We fine-tune the official NeuroBOLT implementation (Li et al., 2024) from the released LaBraM-base weights with the published recipe (AdamW, learning rate 3×10^{-4}, weight decay 0.05, layer decay 0.65, five warm-up epochs followed by a cosine schedule, 20 epochs, batch size 16, drop-path 0.1, MSE loss) and report the last epoch under three seeds. Four adaptations are needed. (1) LaBraM indexes its channel embedding by 10–20 name, and HBN uses a 128-channel EGI HydroCel net, so we feed the 32 electrodes closest to 32 standard 10–10 sites (all within 21 mm of the nominal position). (2) The HBN cache stores min-max scaled EEG, so each channel is z-scored over the 30-second window and scaled by 0.2 to approximate LaBraM's 0.1 mV unit. Unit scale gives the same validation loss and a similar result (0.420 ± 0.046 subject-generalization RSA-ROI over three

seeds). (3) NeuroBOLT reads 16-second windows, so training draws a random 16-second crop of the 30-second HBN window and evaluation averages the predictions of three crops. (4) HBN pairs one EEG window with a 16-frame series of 64 regions, so one regression head predicts all 16×64 standardized values in place of one region per model. Scores are de-standardized before RSA-ROI is computed on the same regional series as for the ROI predictor. Training from scratch gives 0.389 on subject generalization, and the best-validation checkpoint of the published protocol selects epochs 8 to 20 and scores lower and less consistently (0.229 ± 0.153).

C Implementation details

C.1 Data and latent representation

Every experiment predicts a $T = 16$-frame fMRI window from one EEG segment and the subject's sMRI. The cohorts differ in EEG length and repetition time. (1) HBN uses 128-channel EEG sampled at 200 Hz, a 30-second segment ($L = 6000$), and a 12.8-second fMRI window at TR $= 0.8$ s, acquired in separate sessions. (2) The simultaneous cohorts use 63-channel, trigger-aligned EEG sampled at 200 Hz, with a 32-second window ($L = 6400$) at TR $= 2.0$ s for ds002336 and a 16-second window ($L = 3200$) at TR $= 1.0$ s for ds002338. Joint training (`multi`) crops the EEG of ds002336 to 16 seconds. The frozen 3-D VAE maps each volume to a latent of size $(C, H, W, D) = (32, 10, 10, 6)$.

C.2 Orthogonality of the target decomposition

For any time-invariant tensor $\mathbf{A}$ broadcast over T frames and $\mathbf{R}^\star = \mathrm{P}_0(\mathbf{Z}_1)$,

$$\langle \mathbf{A}, \mathbf{R}^\star \rangle = \sum_{r=1}^{T} \langle \mathbf{A}, \mathbf{R}^\star_{:,r} \rangle = T \, \langle \mathbf{A}, \mathrm{mean}_T(\mathbf{R}^\star) \rangle = 0. \tag{13}$$

Thus, $\mathcal{L}_{\mathrm{static}}$ and the residual losses act on orthogonal subspaces, and the anatomy branch receives no gradient from the temporal fluctuations. Within the residual subspace, $\mathbf{D}$ and $\mathbf{C}_\theta$ are separated by subtraction. $\mathbf{D}$ is supervised against $\mathbf{R}^\star$, and $\mathbf{C}_\theta$ fits the stop-gradient remainder $\mathbf{C}^\star = \mathbf{R}^\star - \mathrm{sg}(\mathbf{D})$, so $\mathbf{C}_\theta$ is supervised only on what $\mathbf{D}$ leaves unexplained.

C.3 VibeSync encoder

Input stem. The EEG is average-pooled with window and stride 10, which reduces the sampling rate to 20 Hz and yields $T_e = L/10$ steps (600, 640, and 320 for the 30-, 32-, and 16-second windows). At each step, a two-layer projection with layer normalization and LeakyReLU maps the EEG to the oscillator drive $\boldsymbol{\gamma}^k \in \mathbb{R}^N$. Dropout is disabled so that the encoder is deterministic, as zero-referencing requires.

Electrode graph and phase coupling. The electrode affinity is $\mathbf{A}_e = [\mathbf{P} + \mathbf{P}^\top]_+ + \delta$ with $\delta = 10^{-6}$ and identity-initialized $\mathbf{P}$. Each phase state has $d_\theta = 4$ dimensions and is initialized as a normalized constant vector. $\boldsymbol{\Omega}$ consists of two learnable planar rotations on paired phase dimensions, and Equation (4) uses $\Delta t = 0.1$.

Spiking integrator. The gate delay is $\tau = 2$, and the first step is used for $k \le \tau$. Each electrode maintains $B_r = 4$ dendritic branches $\mathbf{d}^k \in \mathbb{R}^{N \times B_r}$, and the $\mathrm{LIF}(\cdot)$ update in Equation (5) is

$$\begin{aligned} \mathbf{d}^k &= \boldsymbol{\eta} \odot \mathbf{d}^{k-1} + (1 - \boldsymbol{\eta}) \odot \mathrm{reshape}\left(\mathbf{W}_d[\mathbf{r}^k \,\|\, \mathbf{z}^{k-1}] + \mathbf{b}_d\right), \\ \mathrm{LIF}(\cdot) &= \mathbf{a} \odot \mathbf{u}^{k-1} + (1 - \mathbf{a}) \odot \textstyle\sum_{b=1}^{B_r} \mathbf{d}^k_{:,b} - v_{\mathrm{th}} \mathbf{z}^{k-1}, \end{aligned} \tag{14}$$

where $\boldsymbol{\eta}$ and $\mathbf{a}$ are learnable decays in $(0, 1)$. The forward pass uses the hard threshold, and back-propagation uses a multi-Gaussian surrogate derivative.

Tokenizer and zero-referencing. The spiking readout layer maps $\mathbf{z}^k$ to membrane features $\mathbf{h}^k \in \mathbb{R}^N$, which are aggregated as $\mathbf{F}_E = \mathrm{LN}\left(\mathbf{W}_o \, \mathrm{AAP}_T(\mathrm{Conv1D}_5([\mathbf{h}^1, \ldots, \mathbf{h}^{T_e}]))\right) \in \mathbb{R}^{T \times d_e}$, where $\mathrm{Conv1D}_5$ is a temporal convolution with kernel size 5 and AAP_T is adaptive average pooling to T bins. Since the recurrent states are reinitialized at every forward pass, the all-zero reference is concatenated with $\mathbf{E}$ in the same batch, so that both terms of the zero-referencing share one state initialization.

C.4 EEG RESIDUAL, ANATOMY BRANCH, AND CORRECTION FLOW

EEG residual D. The temporal mixer is $f_{\text{mix}}(\widehat{\mathbf{F}}_E) = \mathbf{V} + \text{Conv1D}_3^{\times 2}(\mathbf{V})$ with $\mathbf{V} = \text{SiLU}(\mathbf{W}_h\,\text{LN}(\widehat{\mathbf{F}}_E)) \in \mathbb{R}^{T\times d_h}$, where $\text{Conv1D}_3^{\times 2}$ denotes two temporal convolutions with kernel size 3. We use $K = 16$ spatial modes on HBN and $K = 12$ on the simultaneous cohorts, with a fixed scale $\kappa = 0.1$ and $d_h = d_e$.

Anatomy branch. f_A consists of an input $3\times3\times3$ convolution, three pre-activation residual blocks (GroupNorm–SiLU–convolution), and an output GroupNorm–SiLU–convolution.

Correction flow. At frame r, the drive of $\mathbf{C}_\theta$ is $\boldsymbol{\gamma}_C^r = \mathbf{W}_\gamma\,\text{AvgPool}_{H,W,D}(\mathbf{x}_{t,r}^C) + \text{MLP}(t) \in \mathbb{R}^{N_C}$ with $N_C = 64$, where $\text{MLP}(t)$ is an MLP on a sinusoidal flow-time embedding, shared across frames. Equations (4) and (5) are applied over the N_C nodes and T frames, and the RMS-normalized basis $\{\widetilde{\mathbf{B}}_n\}$ is separate from the modes of $\mathbf{D}$.

C.5 TRAINING OBJECTIVE AND SAMPLING

Losses. $\text{NMSE}(\mathbf{X},\mathbf{Y}) = \|\mathbf{X}-\mathbf{Y}\|_2^2/\|\mathbf{Y}\|_2^2$ is computed per sample and averaged over the batch. In $\mathcal{L}_{\text{flow}}$, the denominator $1-t$ is clamped to at least 10^{-3}, and $\mathcal{L}_{\text{flow}}$ is disabled for the first three epochs. The regularizer is $\mathcal{L}_{\text{reg}} = \mathcal{L}_{\text{orth}} + 0.1\,\mathcal{L}_{\text{smooth}}$ with

$$\mathcal{L}_{\text{orth}} = \frac{1}{K(K-1)}\sum_{k\neq k'}\cos^2\big(\text{vec}(\mathbf{S}_k),\text{vec}(\mathbf{S}_{k'})\big), \qquad \mathcal{L}_{\text{smooth}} = \mathbb{E}\big\|\mathbf{D}_{:,r+1}-2\mathbf{D}_{:,r}+\mathbf{D}_{:,r-1}\big\|_2^2, \tag{15}$$

where $\mathcal{L}_{\text{orth}}$ discourages redundant spatial modes and $\mathcal{L}_{\text{smooth}}$ suppresses high-frequency fluctuations that would not survive the hemodynamic response.

Auxiliary loss. The simultaneous runs add one term with weight 0.1. The connectivity-preservation loss pools the composed prediction and the target latent over $2\times2\times2$ blocks and over channels, which gives $R = 75$ pseudo-region time courses, and matches their $R\times R$ cosine-similarity matrices $\widehat{\mathbf{G}}$ and $\mathbf{G}$ off the diagonal.

$$\mathcal{L}_{\text{fc}} = \frac{1}{R(R-1)}\sum_{i\neq j}\big(\widehat{G}_{ij}-G_{ij}\big)^2, \tag{16}$$

Sampler. Starting from $\mathbf{C}^{(0)} = \text{P}_0(\boldsymbol{\epsilon})$, $\mathbf{C}_\theta$ is integrated with $S = 20$ Euler steps at $t_\ell = \ell/S$,

$$\mathbf{C}^{(\ell+1)} = \text{P}_0\left(\mathbf{C}^{(\ell)} + \frac{1}{S}\cdot\frac{\mathbf{C}_\theta(\mathbf{C}^{(\ell)},t_\ell)-\mathbf{C}^{(\ell)}}{1-t_\ell}\right), \qquad \ell = 0,\dots,S-1, \tag{17}$$

and the final prediction is composed via Equation (8). The loop never receives $\mathbf{D}$, which yields the identity of Equation (10).

C.6 HYPERPARAMETERS

Table 4 lists all hyperparameters.

D PERMUTATION TEST FOR EEG SPECIFICITY

The paired statistic of Section 2.4 compares the correct EEG of a window with one within-run and one cross-subject control. To obtain a per-subject significance level, we keep this protocol fixed, i.e., the same anatomy branch, the same shared noise, and one EEG-blind correction trajectory per window reused for every EEG, and score (1) the matched EEG, (2) the EEG of every other non-overlapping window of the same run (within-run null), (3) up to 32 same-task windows of other subjects (cross-subject null), and (4) an all-zero recording. For each window, ρ is the score of Section 2.4 applied to the composed prediction. Since $\mathbf{A}$ and $\mathbf{C}_\theta$ are shared across conditions, differences in ρ arise from $\mathbf{D}$ alone. A subject's statistic is the mean ρ over its windows. Its null distribution is built from 5000 permutations, each of which draws one EEG per window from the corresponding null pool, and the one-sided p-value is $(b+1)/5001$, where b counts them permutations whose mean ρ reaches the observed value.

Table 4: Hyperparameters of *Rhythm Flow*.

Module	Hyperparameter	Value
Data	EEG window / fMRI frames T	16–32 s / 16
VAE	latent size (C, H, W, D)	$(32, 10, 10, 6)$
VibeSync	encoder steps T_e / step size Δt	$L/10$ / 0.1
	phase dimension d_θ / gate delay τ	4 / 2
	dendritic branches B_r	4
	token dimension d_e	256 (HBN), 128 (simult.)
EEG residual $\mathbf{D}$	spatial modes K	16 (HBN), 12 (simult.)
	scale κ / mixer width d_h	0.1 / d_e
Correction flow $\mathbf{C}_\theta$	latent nodes N_C	64
Loss	λ_Δ / λ_E / λ_r / λ_R	0.5 / 0.1 / 0.5 / 10^{-3}
	margin m / temperature β	0.03 / 0.10
	flow warm-up	3 epochs
	auxiliary losses (simult.)	$\mathcal{L}_{\mathrm{fc}}$, weight 0.1
Training	optimizer	AdamW
	learning rate / weight decay / clip	10^{-4} / 10^{-3} / 1.0
	schedule	cosine with warm restarts
	batch size	4 (HBN), 2 (simult.)
	epochs (HBN)	≤ 300
	checkpoint (HBN)	largest EEG-conditioning gap
	epochs (simult.)	≤ 100, early stopping patience 8
Inference	Euler steps S	20

Table 5 lists the results. On ds002336, `full` outperforms both nulls with $p < 0.05$ for both subjects in every seed. On ds002338, the specificity is carried by xp216, whose 36 windows separate from both nulls with $p \leq 0.002$ in every seed of `full` and `multi`, while xp201 and xp206, with 90 windows each, do not separate reliably. The number of windows therefore does not limit detection on this cohort. The table contains 60 tests, so the Bonferroni threshold at a family-wise level of 0.05 is 8.3×10^{-4}. Every entry below 10^{-3} lies between 2×10^{-4}, the smallest attainable value, and 6×10^{-4}, so these tests remain significant under this correction.

Table 5: Permutation test for EEG specificity on the validation subjects. Advantage is the mean within-run advantage $\rho^+ - \bar{\rho}^-_{\mathrm{run}}$ over a subject's windows ($\times 10^{-3}$), listed as seed 0/1/2. p_{run} and p_{cross} are one-sided permutation p-values against the within-run and cross-subject nulls (5000 permutations).

Model	Cohort	Subject	Windows	Advantage	p_{run}	p_{cross}
`full`	ds002336	xp102	22	+1.5/+3.3/+0.7	0.0005/0.0002/0.015	0.0005/0.0002/0.0004
`full`	ds002336	xp109	22	+1.6/+2.3/+1.1	0.0005/0.0002/0.003	0.002/0.0006/0.010
`full`	ds002338	xp201	90	+0.5/+0.3/+0.6	0.140/0.225/0.108	0.070/0.296/0.085
`full`	ds002338	xp206	90	+0.6/+0.2/+0.5	0.018/0.242/0.089	0.100/0.361/0.170
`full`	ds002338	xp216	36	+1.9/+3.1/+3.0	0.0004/0.0002/0.0002	0.002/0.0002/0.0002
`multi`	ds002336	xp102	22	+0.1/+0.0/+0.0	0.198/0.394/0.467	0.123/0.308/0.370
`multi`	ds002336	xp109	22	+0.6/+0.3/+0.3	0.005/0.053/0.082	0.014/0.049/0.152
`multi`	ds002338	xp201	90	+0.1/+0.2/+0.3	0.426/0.320/0.233	0.472/0.417/0.128
`multi`	ds002338	xp206	90	+0.2/+0.2/+0.4	0.242/0.204/0.044	0.289/0.349/0.135
`multi`	ds002338	xp216	36	+3.2/+2.0/+2.0	0.0002/0.0002/0.0002	0.0002/0.0006/0.0002

E SUBJECT CROSS-VALIDATION ON DS002336

We partition the ten ds002336 subjects into five folds of two sealed test subjects, use the next two subjects for checkpoint selection, and train on the remaining six. Each fold is trained once (seed 2027) for `full`, for `no_flow`, which removes $\mathbf{C}_\theta$, and for `no_eeg`, which removes the EEG pathway. `full` and `no_flow` are evaluated on the test subjects at the checkpoint with the largest validation EEG-conditioning gap (`sel.`) and at the last epoch (`final`). The selected epochs range from 2 to 14 for `full` and from 1 to 12 for `no_flow`, and the final checkpoints lie at epochs 16

to 18. Table 6 reports, per fold, the out-of-fold RSA-ROI of matched EEG, the mean within-run advantage over the fold's windows, and the per-subject permutation p-values of Appendix D.

Two observations follow. First, the EEG residual is specific out of fold. Without the correction flow, the within-run advantage is positive for nine of ten test subjects at the selected checkpoint ($+0.78 \times 10^{-3}$, one-sided Wilcoxon $p = 0.024$, sign test $p = 0.011$, cross-subject null $p = 0.032$) and for eight of ten at the final checkpoint ($+0.97 \times 10^{-3}$, $p = 0.042$). With the correction flow, it is positive for eight of ten subjects at both checkpoints ($+0.01 \times 10^{-3}$ and $+0.21 \times 10^{-3}$, $p = 0.065$ and 0.053). Second, the correction flow supplies the representational structure. It raises the mean RSA-ROI of matched EEG from 0.023 to 0.135 at the selected and from 0.084 to 0.174 at the final checkpoint.

Table 6: Five-fold subject cross-validation on ds002336. `sel.` is the checkpoint with the largest validation EEG-conditioning gap and `final` the last epoch. **(a)** Out-of-fold RSA-ROI of matched EEG. **(b)** Mean within-run advantage $\rho^{+} - \bar{\rho}^{-}_{\mathrm{run}}$ over the fold's windows ($\times 10^{-3}$). "Positive" counts the test subjects with a positive advantage, and the Wilcoxon test is one-sided over the ten test subjects. **(c)** One-sided permutation p-values of the two test subjects against the within-run null (5000 permutations), in the order of the test-subject column. `no_eeg` has no EEG residual, so only its RSA-ROI is defined.

		`full`		`no_flow`		`no_eeg`
Fold	**Test subjects**	`sel.`	`final`	`sel.`	`final`	
(a) *Out-of-fold RSA-ROI*↑						
0	xp101, xp102	0.174	0.210	0.000	0.120	−0.040
1	xp103, xp104	0.073	0.209	0.040	0.190	0.247
2	xp105, xp106	0.008	0.019	0.000	0.003	−0.011
3	xp107, xp108	0.102	0.093	0.060	0.083	0.227
4	xp109, xp110	0.319	0.341	0.015	0.024	0.268
Mean		0.135	0.174	0.023	0.084	0.138
(b) *Within-run advantage* ($\times 10^{-3}$)↑						
0	xp101, xp102	−0.64	−0.64	−0.48	−1.03	–
1	xp103, xp104	+0.27	+0.42	+1.04	+1.34	–
2	xp105, xp106	+0.00	+0.32	+0.04	+1.31	–
3	xp107, xp108	+0.40	+0.57	+1.32	+1.18	–
4	xp109, xp110	+0.02	+0.36	+1.99	+2.06	–
Positive		8/10	8/10	9/10	8/10	–
Wilcoxon p		0.065	0.053	0.024	0.042	–
(c) *Per-subject permutation p (within-run null)*↓						
0	xp101, xp102	0.999, 0.040	0.999, 0.041	0.997, 0.260	0.990, 0.099	–
1	xp103, xp104	0.003, 0.013	0.005, 0.099	0.005, 0.007	0.002, 0.039	–
2	xp105, xp106	0.808, 0.132	0.022, 0.149	0.470, 0.425	0.002, 0.281	–
3	xp107, xp108	0.002, 0.200	0.0002, 0.061	0.001, 0.110	0.004, 0.188	–
4	xp109, xp110	0.079, 0.438	0.017, 0.857	0.0002, 0.280	0.0002, 0.846	–