EMAG: DIFFERENTIABLE 4D GAUSSIAN MIXTURE SPLATTING FOR EEG SPATIAL SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

High-density electroencephalography (HD-EEG) enables fine-grained measurement of cortical activity but requires expensive hardware and lengthy setup times, limiting its clinical and research accessibility. We propose **EMAG** (**E**EG **M**ixture of **A**nisotropic **G**aussians), a differentiable framework that reconstructs HD-EEG signals from a sparse subset of low-density (LD) electrodes by representing brain electrical sources as a mixture of anisotropic 4D space-time Gaussians. EMAG places a mixture of multiple Gaussians at each point of a spherical brain grid, each parameterized by a full 4×4 precision matrix, enabling anisotropic spatial spreads and explicit coupling between spatial and temporal dimensions. The forward model renders scalp EEG via differentiable Gaussian field contributions at electrode locations, enabling end-to-end training without explicit source localization supervision. We evaluate EMAG on three public EEG benchmarks (Localize-MI, SEED, and SEED-IV) at super-resolution factors of $2\times$ through $16\times$. EMAG outperforms the current state-of-the-art EEG super-resolution method at most super-resolution factors. The learned Gaussians carry explicit 3D brain-space coordinates and can be directly visualized; this spatial grounding potentially opens avenues for clinical and neuroscientific applications such as source localization.

1 INTRODUCTION

Electroencephalography (EEG) is the most widely used non-invasive neuroimaging modality for measuring cortical electrical activity with millisecond temporal resolution (Ramadan & Vasilakos, 2017; Wolpaw, 2007). While standard clinical EEG employs 19–32 electrodes, high-density (HD) systems with 128–256 channels are increasingly recognized as essential for accurate source localization (Grech et al., 2008) and brain-computer interfaces (Lawhern et al., 2018). And yet, such HD-EEG acquisition remains impractical in many settings due to the cost of electrode caps, extended preparation time, and patient discomfort.

Given this gap, *EEG super-resolution*, the problem of reconstructing HD-EEG from a sparse low-density (LD) subset, has emerged as a practical alternative (Corley & Huang, 2018). Existing approaches treat EEG super-resolution as a channel-to-channel mapping problem, using convolutional, adversarial, or diffusion-based architectures that operate directly in electrode space (Corley & Huang, 2018; Liu et al., 2025a; Wang et al., 2025). None of these methods incorporates an explicit forward model of how brain sources project onto scalp electrodes, forgoing the physical inductive bias that could improve both reconstruction accuracy and interpretability.

Unrelated to EEG analysis, the computer vision community has developed powerful differentiable rendering frameworks, most notably Neural Radiance Fields (NeRF) (Mildenhall et al., 2021) and 3D Gaussian Splatting (Kerbl et al., 2023), that represent volumetric scenes as continuous fields and render them via differentiable forward models. Recent extensions to 4D Gaussian Splatting (Duan et al., 2024; Wu et al., 2024) further incorporate temporal dynamics via deformation fields or native 4D covariance matrices.

Recent work on Neural Brain Fields (NBF) (Kedem et al., 2025) has drawn on the NeRF analogy for EEG, using implicit neural representations to generate signals at unseen electrode positions. However, NBF uses a per-sample Multilayer Perceptron (MLP) to encode the electromagnetic potential as a continuous function of 3D position and time, without explicit source modeling, and remains

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

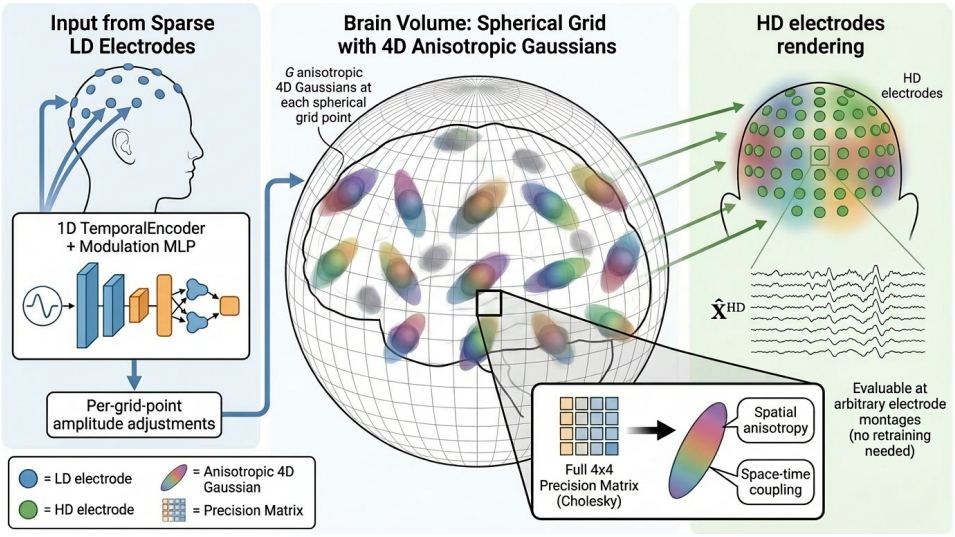

Figure 1: Overview of EMAG. (1) LD electrodes drive a 1D TemporalEncoder + MLP that outputs amplitude adjustments. (2) The brain volume hosts G anisotropic 4D Gaussians per spherical-grid point. (3) Differentiable rendering at electrode positions yields the HD scalp signal $\hat{\mathbf{X}}^{\mathrm{HD}}$.

fundamentally non-interpretable in terms of neural sources, as it offers no mechanism to localize or characterize the underlying brain sources.

While these two fields have been progressing independently, we observe a structural analogy between volumetric scene rendering and the EEG forward problem: both involve a latent volumetric field (radiance vs. source current density) that is projected onto a set of sensors (camera pixels vs. scalp electrodes). This motivates our approach, but requires adaptation: while 4D Gaussian Splatting (Wu et al., 2024) handles dynamic scenes via space-time covariance, it lacks a forward model for EEG's volume-conduction physics (leadfield source-to-scalp projection). We thus extend it with anisotropic 4D mixtures on a spherical grid and differentiable Gaussian rendering to scalp potentials: a physics-inspired surrogate that encodes the core volume-conduction inductive bias, a smooth distance-decaying source-to-sensor projection, rather than an exact biophysical head model. This enables end-to-end SR without explicit inverse supervision—outperforming baselines on most settings.

Contributions. We introduce **EMAG** (**E**EG **M**ixture of **A**nisotropic **G**aussians), a framework for EEG super-resolution that advances the application of Gaussian splatting to the neuroscience domain (Fig. 1):

1. We formulate EEG super-resolution as **differentiable rendering of a 4D source field**, where the "scene" is modeled as a *mixture* of anisotropic Gaussians placed on a spherical brain grid and the "rendering" is a weighted summation of gaussians at electrode positions.
2. We propose an **anisotropic 4D Gaussian parameterization** in which each component is equipped with a full space-time 4×4 precision matrix parameterized via Cholesky decomposition, enabling anisotropic spatial spreads and explicit space-time coupling.
3. We provide a **systematic evaluation** on three public EEG benchmarks (Localize-MI (Mikulan et al., 2020), SEED (Duan et al., 2013; Zheng & Lu, 2015), and SEED-IV (Zheng et al., 2018)) across super-resolution factors of $2\times$ through $16\times$, including comparison against seven established EEG-SR baselines.

2 Related Work

EEG source reconstruction. The EEG inverse problem, recovering brain source activity from scalp measurements, is fundamentally ill-posed since the mapping from source to sensor space is many-to-one (Grech et al., 2008; Hallez et al., 2007). Like many sampling scenarios, it becomes

more severe with fewer or sparser samples. Classical methods such as minimum norm estimation (MNE) (Dale et al., 2000; Hämäläinen & Ilmoniemi, 1994) and standardized low-resolution brain electromagnetic tomography (sLORETA) (Pascual-Marqui et al., 2002) impose regularization priors (e.g., ℓ_2 smoothness) to select among infinitely many solutions. These methods require an explicit forward model (leadfield matrix) derived from structural MRI and head tissue conductivity models (Hämäläinen & Ilmoniemi, 1994; Michel & Brunet, 2019). Recent work has explored data-driven or deep learning–based EEG source imaging that learns the source-to-sensor mapping end-to-end rather than relying solely on a fixed leadfield (Huang et al., 2022; Jiao et al., 2024; Sun et al., 2022). Our approach follows this trend but differs by parameterizing sources as explicit anisotropic 4D Gaussians on an anatomically anchored brain grid and embedding the forward operator as a differentiable Gaussian rendering step.

EEG super-resolution. Classical EEG channel interpolation relied on spatial methods rather than learned super-resolution, most commonly spherical-spline interpolation (Perrin et al., 1989); these interpolate scalp potentials directly and do not model the underlying neural sources. Deep learning approaches to EEG channel interpolation include convolutional and adversarial architectures that map LD to HD signals directly (Corley & Huang, 2018; Liu et al., 2023; Pérez et al., 2025; Saba-Sadiya et al., 2020). The Step-Aware Residual Guide (Liu et al., 2025a) introduces residual learning with temporal step-aware conditioning, achieving strong performance with a lightweight CNN backbone by exploiting temporal structure. While effective, these methods do not model the 3D geometry of the brain and electrodes, limiting physical interpretability. Our method explicitly models sources in 3D brain space and renders to arbitrary electrode configurations.

Neural fields and Gaussian splatting. Neural Radiance Fields (NeRF) (Mildenhall et al., 2021) represent scenes as continuous functions mapping spatial coordinates to radiance, using sinusoidal positional encodings (Tancik et al., 2020; Vaswani et al., 2017) to capture high-frequency detail. 3D Gaussian Splatting (Kerbl et al., 2023) replaces the MLP with an explicit set of anisotropic 3D Gaussians (each with a full 3×3 covariance matrix), achieving real-time rendering. 4D extensions incorporate temporal dynamics: Wu et al. (2024) combine 3D Gaussians with deformation fields, while 4D-Rotor Gaussian Splatting (Duan et al., 2024) uses native 4×4 spatiotemporal covariance matrices.

Neural representations for EEG. Neural Brain Fields (NBF) (Kedem et al., 2025) draws on the NeRF analogy for EEG, training a per-sample implicit neural representation to generate signals at unseen electrode positions. While NBF demonstrates the potential of neural fields for EEG interpolation, it uses a black-box MLP without explicit source modeling or Gaussian parameterization. Our approach extends this line of work by introducing explicit Gaussian source components with full 4D covariance, combined with a physics-inspired forward model.

3 METHOD

3.1 PROBLEM FORMULATION

Let $\mathbf{X}^{\text{HD}} \in \mathbb{R}^{M \times T}$ denote an HD-EEG recording with M electrodes and T time steps, and let $\mathbf{X}^{\text{LD}} \in \mathbb{R}^{m \times T}$ ($m < M$) be a low-density subset obtained by selecting m of the M channels. The super-resolution factor is $r = M/m$. Our goal is to learn a mapping $f_\theta : \mathbb{R}^{m \times T} \to \mathbb{R}^{M \times T}$ that reconstructs $\hat{\mathbf{X}}^{\text{HD}} \approx \mathbf{X}^{\text{HD}}$ from $\mathbf{X}^{\text{LD}}$ alone.

Rather than learning this mapping directly in electrode space, we decompose it into two stages: (1) infer a latent 4D source field $\mathbf{S}(\mathbf{x}, t)$ for $\mathbf{x} \in \mathbb{R}^3$ and $t \in \{1, \ldots, T\}$, conditioned on the LD input; and (2) render the HD-EEG via a differentiable forward model that maps $\mathbf{S}$ to scalp potentials.

Our explicit Gaussian representation of $\mathbf{S}$ enables rendering to arbitrary HD electrode montages of any size, without retraining.

3.2 FORWARD MODEL: ANISOTROPIC 4D GAUSSIAN FIELD RENDERING

Gaussian source components are anchored on a discrete brain grid $\mathcal{G} = \{\mathbf{g}_i\}_{i=1}^{N}$. We generate candidate points on a uniform cubic lattice spanning $[-90, 90]$ mm in each spatial dimension at

resolution R (default $R = 12$) and retain only those within a sphere of radius $r_{\text{brain}} = 90$ mm:

$$\mathcal{G} = \{\mathbf{g}_i : \|\mathbf{g}_i\| \leq r_{\text{brain}}\}. \tag{1}$$

This yields $N \approx 672$ (at $R = 12$) grid points, constraining sources to the anatomically relevant brain volume while keeping the grid coarse enough for efficient rendering.

At each grid point $\mathbf{g}_i$, we place G Gaussian components (default $G = 3$), for a total of $N_{\text{total}} = N \cdot G$ components. Importantly, seeking a realistic setup, we assume time-dependent activation and thus allow time-modulated Gaussian components. The predicted EEG signal at electrode j and time t is computed as a sum of Gaussian field contributions from all source components:

$$\hat{X}_{j,t} = \sum_{n=1}^{N_{\text{total}}} a_n(t) \cdot \exp\left(-\frac{1}{2}\,\mathbf{d}_{j,n,t}^{\top}\boldsymbol{\Sigma}_n^{-1}\mathbf{d}_{j,n,t}\right), \tag{2}$$

where $a_n(t)$ is the modulated amplitude of component n at time t (Section 3.4), $\mathbf{d}_{j,n,t} = [\mathbf{e}_j - \mathbf{g}_n;\, t_{\text{norm}} - \mu_n^{(t)}] \in \mathbb{R}^4$ is the 4D displacement vector between electrode j and component n at time t, and $\boldsymbol{\Sigma}_n^{-1}$ is the 4×4 precision matrix of the n-th Gaussian.

This formulation is analogous to Gaussian splatting in computer vision (Kerbl et al., 2023): each source "splats" its contribution to nearby electrodes with a weight that decays as an anisotropic Gaussian function of the 4D space-time distance.

3.3 Mixture of anisotropic 4D Gaussians

The key innovation of EMAG is twofold: (1) placing *multiple* Gaussian components at each grid point to form a mixture, enabling complex multi-modal source patterns, and (2) parameterizing each component with a full 4×4 precision matrix, enabling anisotropic spatial fields and explicit space-time coupling.

Cholesky parameterization. A precision matrix must be symmetric positive-definite (SPD) to define a valid Gaussian shape. We ensure this by construction via the Cholesky decomposition: we parameterize $\boldsymbol{\Sigma}_n^{-1}$ as $\mathbf{L}_n\mathbf{L}_n^{\top}$, where $\mathbf{L}_n$ is a lower-triangular matrix with positive diagonal entries. Since any product $\mathbf{L}\mathbf{L}^{\top}$ is automatically SPD, the network can learn unconstrained values for $\mathbf{L}$ and the result is always a valid precision matrix:

$$\boldsymbol{\Sigma}_n^{-1} = \mathbf{L}_n\mathbf{L}_n^{\top}, \quad \mathbf{L}_n = \begin{bmatrix} d_0 & 0 & 0 & 0 \\ c_0 & d_1 & 0 & 0 \\ c_1 & c_2 & d_2 & 0 \\ c_3 & c_4 & c_5 & d_3 \end{bmatrix}, \tag{3}$$

where $d_i = \exp(\ell_i)$ are positive diagonal entries (stored as log-values ℓ_i for unconstrained optimization) and c_j are free off-diagonal entries. The diagonal entries control the precision (inverse of spread) along each axis: d_0, d_1, d_2 for the three spatial dimensions and d_3 for time, while the off-diagonal entries $c_0, \ldots, c_5$ couple the dimensions: c_0, c_1, c_2 orient the spatial ellipsoid off-axis, while c_3, c_4, c_5 let a component's spatial shape vary over time.

Each of the N_{total} Gaussian components thus carries 12 learnable scalars: a signed **base amplitude** w_n (overall source strength); a **temporal center** $\mu_n^{(t)} \in [0, 1]$ (time of peak activation); the **Cholesky log-diagonal** $(\ell_0, \ell_1, \ell_2, \ell_3)$ controlling per-axis precision; and the **Cholesky off-diagonals** $(c_0, \ldots, c_5)$ encoding spatial anisotropy and space-time coupling.

Efficient Mahalanobis decomposition. The exponent in Eq. 2 requires computing the Mahalanobis distance $\mathbf{d}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{d} = \|\mathbf{L}^{\top}\mathbf{d}\|^2$, which measures the distance from an electrode to a Gaussian center through the Gaussian's precision matrix. Naïvely computing this for every (electrode, component, timestep) triple requires materializing a $(T, M, N_{\text{total}}, 4)$ tensor, which is computationally expensive. We exploit the fact that electrode and grid positions are fixed across time to decompose the distance into precomputable spatial terms and cheap per-timestep temporal terms:

$$\|\mathbf{L}^{\top}\mathbf{d}\|^2 = \underbrace{A(\mathbf{e}_j, \mathbf{g}_n)}_{\text{spatial only}} + 2\tau\,\underbrace{C(\mathbf{e}_j, \mathbf{g}_n)}_{\text{coupling}} + \tau^2\,\underbrace{D_n}_{\text{temporal only}}, \tag{4}$$

where $\tau = t_{\text{norm}} - \mu_n^{(t)}$ and:

$$z_0 = d_0\Delta x + c_0\Delta y + c_1\Delta z, \quad z_1 = d_1\Delta y + c_2\Delta z, \quad z_2 = d_2\Delta z, \tag{5}$$

$$A = z_0^2 + z_1^2 + z_2^2, \quad C = z_0c_3 + z_1c_4 + z_2c_5, \quad D = c_3^2 + c_4^2 + c_5^2 + d_3^2. \tag{6}$$

The terms A, C, and D are **precomputed once** per forward pass for all electrode–component pairs by combining the fixed spatial offsets $(\Delta x, \Delta y, \Delta z)$ with the current Cholesky parameters (d_i, c_j) of each component. Within a forward pass these quantities stay constant across time, while only the temporal displacement τ changes, so at each timestep we obtain the Mahalanobis distance by evaluating $\|\mathbf{L}^\top \mathbf{d}\|^2 = A(\mathbf{e}_j, \mathbf{g}_n) + 2\tau C(\mathbf{e}_j, \mathbf{g}_n) + \tau^2 D_n$, rather than recomputing a full 4D product for every time step.

3.4 Low-density conditioning

Rather than predicting source activity directly from space–time coordinates, we use the LD signals to drive a low-dimensional temporal code that modulates the amplitudes of Gaussians anchored on the brain grid, so that spatial structure is encoded in the grid while per-sample dynamics are driven only by the observations. To condition the source field on the observed LD-EEG, we then need a mechanism that tells the Gaussian sources "how active to be" at each time step based on the sparse input. We extract per-timestep features from $\mathbf{X}^{\text{LD}}$ using a lightweight 1D temporal encoder:

$$\mathbf{h}_t = \text{TemporalEncoder}(\mathbf{X}^{\text{LD}}_{\cdot,t}) \in \mathbb{R}^{d_f}, \tag{7}$$

where d_f is the feature dimension. This encoder reads the LD electrode values at time t and compresses them into a feature vector. A modulation network then maps this feature into a per-grid-point amplitude adjustment:

$$\Delta w_i(t) = \text{MLP}_{\text{mod}}(\mathbf{h}_t)_i, \quad i = 1, \ldots, N. \tag{8}$$

The modulation is shared across all G Gaussians at the same grid point, preserving the structural relationship between co-located components. The final amplitude for the g-th Gaussian at grid point i at time t is:

$$a_{i,g}(t) = w_{i,g} + \Delta w_i(t). \tag{9}$$

In other words, each Gaussian has a fixed base amplitude $w_{i,g}$ (learned during training) plus a time-varying adjustment $\Delta w_i(t)$ driven by the LD input. This separates the *static* source layout (base amplitudes, spatial-temporal covariance) from the *dynamic* modulation driven by the observed signal, analogous to how 4D Gaussian splatting separates static geometry from per-frame deformations (Wu et al., 2024).

3.5 Training objective

The model is trained end-to-end by minimizing the mean squared error between predicted and ground-truth HD-EEG, plus a small L_2 penalty on the source covariance and amplitude parameters that keeps the precision matrices well-conditioned:

$$\mathcal{L} = \frac{1}{M \cdot T} \sum_{j=1}^{M} \sum_{t=1}^{T} \big(\hat{X}_{j,t} - X^{\text{HD}}_{j,t}\big)^2 + \lambda \sum_{\theta \in \Theta_{\text{reg}}} \frac{1}{|\theta|} \|\theta\|_2^2, \tag{10}$$

where $\Theta_{\text{reg}} = \{\boldsymbol{\ell}_n, \mathbf{c}_n, w_n\}_n$ collects the Cholesky log-diagonal, off-diagonal, and base-amplitude parameters of every Gaussian, and $\lambda = 2 \times 10^{-4}$.

4 Experimental evaluations

4.1 Datasets

We evaluate EMAG on three public EEG benchmarks that span different experimental paradigms and are standard in the EEG super-resolution literature (preprocessing details in App. B):

Localize-MI. The Localize-MI dataset (Mikulan et al., 2020) provides 256-channel HD-EEG recordings from 7 human subjects during intracerebral electrical stimulation. Intracranial electrodes are

Table 1: EEG super-resolution results across three datasets: **Localize-MI** (256 channels), **SEED**, and **SEED-IV** (62 channels). Best in **bold**, second-best underlined. Metrics: NMSE (↓), PCC (↑), SNR (↑, dB). EMAG numbers are mean over five seeds (±std in App. I); baseline results are as reported by Liu et al. (2025a).

Model	Metric	Localize-MI (256 ch.)				SEED (62 ch.)			SEED-IV (62 ch.)		
		2×	4×	8×	16×	2×	4×	8×	2×	4×	8×
SaSDim	NMSE	0.2675	0.3427	0.4174	0.4613	0.4399	0.6234	0.7767	0.3633	0.5543	0.7122
	PCC	0.8194	0.7246	0.6926	0.6476	0.7341	0.5649	0.4349	0.7249	0.6211	0.5009
	SNR	5.7443	4.3796	3.5549	2.7678	4.1154	2.2940	1.1349	4.5940	2.6004	1.6211
SADI	NMSE	0.2637	0.3442	0.4164	0.4566	0.4439	0.6049	0.8106	0.3557	0.5349	0.6844
	PCC	0.8243	0.7391	0.6944	0.6554	0.7234	0.5819	0.4064	0.7624	0.6293	0.5243
	SNR	5.7511	4.3724	3.5498	2.8942	4.2419	2.5160	1.0137	4.7093	2.6044	1.6610
RDPI	NMSE	0.2561	0.3562	0.4076	0.4531	0.4064	0.6134	0.7916	0.3491	0.5416	0.6915
	PCC	0.8246	0.7396	0.7062	0.6549	0.7416	0.5716	0.4216	0.7861	0.6316	0.5164
	SNR	5.7311	4.3966	3.5643	2.7731	4.2619	2.3160	1.0316	4.7190	2.6194	1.6492
DDPMEEG	NMSE	0.2046	0.3108	0.3554	0.4076	0.4916	0.7319	0.8634	0.5136	0.6513	0.7916
	PCC	0.8516	0.8163	0.7306	0.6739	0.6941	0.5134	0.3419	0.7346	0.5316	0.4305
	SNR	6.2151	5.5126	3.9891	3.2715	4.1943	1.5391	0.9431	4.4165	2.1064	1.6105
ESTformer	NMSE	0.2721	0.3578	0.4466	0.4837	0.3288	0.3483	0.4149	0.3448	0.3911	0.5125
	PCC	0.8061	0.7205	0.6867	0.6319	0.8368	0.8012	0.7670	0.8106	0.7822	0.7048
	SNR	5.5403	3.8671	3.3023	2.5671	5.0560	4.5838	3.8871	4.7535	4.1933	2.9821
STAD	NMSE	0.1902	0.3067	0.3649	0.4106	0.4319	0.6913	0.8671	0.3819	0.6713	0.7193
	PCC	0.8635	0.8194	0.7216	0.6694	0.7136	0.4946	0.3441	0.7316	0.5219	0.4319
	SNR	7.2591	5.5234	3.8715	3.2642	4.1364	1.4349	0.9134	4.4930	2.0492	1.6193
SRGDiff	NMSE	0.1449	0.2384	0.2957	0.3457	0.1632	0.2977	**0.3494**	0.1663	0.2115	**0.2603**
	PCC	0.9213	0.8854	0.8323	0.7322	0.9102	0.8445	**0.8167**	0.9113	0.8846	**0.8210**
	SNR	8.3755	6.3617	5.2249	4.0197	7.8413	5.2606	**4.5912**	7.8660	6.6402	**6.0346**
EMAG (ours)	NMSE	**0.0856**	**0.1040**	**0.1396**	**0.2139**	**0.1545**	**0.2776**	0.4027	**0.1153**	**0.2027**	0.3358
	PCC	**0.9561**	**0.9464**	**0.9274**	**0.8863**	**0.9194**	**0.8496**	0.7719	**0.9404**	**0.8926**	0.8128
	SNR	**10.8996**	**10.0156**	**8.6731**	**6.7959**	**8.2044**	**5.6243**	3.9989	**9.5780**	**7.0593**	4.7712

used for stimulation only (not measurement); the scalp HD-EEG captures the brain's response. The known stimulation sites provide approximate ground-truth source locations. Each subject has approximately 74 stimulation epochs at a sampling rate of 8 kHz. Given $M = 256$ HD electrodes, we construct LD subsets by uniformly subsampling channels at factors $r \in \{2, 4, 8, 16\}$, yielding $m \in \{128, 64, 32, 16\}$ LD electrodes respectively.

SEED. The SEED dataset (Duan et al., 2013; Zheng & Lu, 2015) contains 62-channel EEG (international 10–20 montage, 200 Hz) recordings from 15 subjects during emotion elicitation via film clips, with three emotion classes (positive, neutral, negative). We use the preprocessed band-passed (0–75 Hz) signals released with the dataset. Each trial is segmented into non-overlapping 4 s windows (800 samples). LD subsets are constructed by uniformly subsampling channels at factors $r \in \{2, 4, 8\}$, yielding $m \in \{31, 15, 7\}$ LD electrodes.

SEED-IV. The SEED-IV dataset (Zheng et al., 2018) extends SEED to four emotion classes (happy, sad, fear, neutral), with the same 62-channel 200 Hz montage across 15 subjects. SR setup, train/test split, and channel-subsampling protocol are identical to SEED ($r \in \{2, 4, 8\}$, $m \in \{31, 15, 7\}$).

For each dataset we use a per-subject 80/20 trial split and report metrics averaged across all of its subjects; LD subset remains fixed per subject across sessions.

4.2 Main results

We compare EMAG against SRGDiff (Liu et al., 2025a)—the current SOTA for deep EEG super-resolution—plus all baselines reported there: SaSDim (Zhang et al., 2023), SADI (Islam et al., 2025), RDPI (Liu et al., 2025b), DDPMEEG (Vetter et al., 2024), ESTformer (Li et al., 2023), and STAD (Wang et al., 2025). Ablations are reported in App. F; per-subject numbers in App. D.3.

Table 1 reports reconstruction quality on Localize-MI (256 ch., 7 subjects), SEED, and SEED-IV (62 ch., 15 subjects each) across all SR factors. EMAG leads on Localize-MI at every r, cutting NMSE by $41-62\%$ and improving SNR by $2.5-4.4$ dB relative to the strongest prior baseline (SRGDiff), while keeping PCC above 0.88 even at $r{=}16$. On SEED and SEED-IV it dominates at $r{\in}\{2, 4\}$

across all three metrics; at $r{=}8$ it trails SRGDiff in NMSE/SNR while remaining second-best in PCC, an outcome consistent with information-theoretic limits of $8\times$ on a 62-channel cap.

Three observations clarify these trends. *(i)* The advantage is largest on Localize-MI, whose dense 256-channel HD-EEG montage rewards EMAG's anisotropic 4D source representation; degradation with r is also slowest there ($\Delta\mathrm{PCC}{=}0.07$ between $r{=}2$ and $r{=}16$, vs. $0.13-0.15$ for the next-best method), indicating stronger robustness to extreme decimation than electrode-space diffusion. *(ii)* On 62-channel scalp data, the EMAG–SRGDiff gap shrinks with r as the inverse problem becomes under-determined, aligning with the LD conditioning ablation (Sec. 4.4) that identifies input spatial resolution as the limiting factor beyond the cap's bandwidth. *(iii)* EMAG achieves this with $\sim$73k parameters per subject—about $30\times$ fewer than SRGDiff ($\sim$2.3M)—and trains in under 30 minutes on a single GPU, supporting its role as a lightweight *personalized* model that nonetheless generalizes across subjects (Sec. 4.3).

Beyond reconstruction fidelity, the gain is meaningful downstream: used as input to a SEED emotion classifier, EMAG's reconstruction improves classification accuracy over the sparse LD input by $+0.20$, $+0.21$, and $+0.17$ at $2\times$ / $4\times$ / $8\times$ respectively, positive for all $15/15$ subjects (App. C).

4.3 Cross-subject generalization

EMAG is by design a *personalized*, per-subject model: the 4D Gaussian field is fit on each subject, and the main results table reports this per-subject regime, which is its strongest. We nonetheless test how far a single model generalizes across subjects. Training one model across subjects, either pooled (fit jointly on all subjects) or leave-one-subject-out (LOSO, evaluated on a fully held-out subject), reconstructs an unseen subject within $\approx$0.03 NMSE of the per-subject fit on average, and on Localize-MI a completely unseen subject is still reconstructed better than the prior SOTA SRGDiff. A single model can therefore serve multiple subjects, including new ones recorded with the same electrode montage. Full pooled and LOSO results are in App. D.2.

As a complementary probe, we also measure single-to-single transfer between individual subjects, comparing against a non-learned spherical-spline upsampling (*LD-spline*) reference; the mean off-diagonal transfer scores and full $S \times S$ matrices are reported in App. D.

4.4 Ablation studies

We conduct a systematic ablation study to validate each core design decision in EMAG, evaluated on Localize-MI and SEED at $4\times$ super-resolution. Full results are reported in Table 2 and a detailed per-ablation discussion is provided in Appendix F.

4D Gaussian parameterization. Reducing the precision matrix from full $4{\times}4$ to spatial-only ($3{\times}3$), diagonal, and isotropic-scalar variants degrades performance monotonically (isotropic: $+19.0\%$ NMSE on Localize-MI, $+22.1\%$ on SEED). Zeroing only the spatial off-diagonals c_0, c_1, c_2 or only the spatio-temporal c_3, c_4, c_5 produces comparable degradation, indicating the two coupling types are complementary and jointly contribute to optimal performance.

Brain grid and source placement. Without the spherical grid anchor, free-floating components migrate to the scalp and act as shortcut interpolators, so the constraint is a geometric regularizer rather than a convenience. Restricting the grid to a thin cortical shell costs $+116.2\%$ NMSE on SEED, challenging the assumption that emotion-related EEG is purely cortical.

Low-density conditioning. Removing the TemporalEncoder collapses NMSE to ≈ 1.0 on both datasets, and replacing per-grid-point modulation with a global scalar costs $+434\%$ NMSE on Localize-MI yet only $+7.6\%$ on SEED, reflecting the spatial focality of evoked responses. Pre-interpolating the LD input yields no gain, indicating the Gaussian forward model already supplies the spatial regularization.

Physics-inspired forward model. A direct electrode-space MLP raises Localize-MI NMSE by $+113.7\%$ but SEED by only $+2.0\%$, again tracking spatial focality. A learned linear source-to-electrode projection matches the MLP on Localize-MI yet collapses on SEED ($+94.0\%$), exposing its lack of a distance-decay prior; the fixed three-shell leadfield is the weakest variant on both datasets, unable to adapt to subject anatomy.

Table 2: Ablation results on **Localize-MI** and **SEED** at $4\times$ super-resolution.

Category	Variant	**Localize-MI** (mean across 7 subjects)			**SEED** (mean across 15 subjects)		
		NMSE ↓	**PCC** ↑	**SNR** ↑	**NMSE** ↓	**PCC** ↑	**SNR** ↑
Proposed Model	**EMAG (full)**	**0.1048**	**0.9460**	**9.9720**	**0.2802**	**0.8480**	**5.5816**
4D Gaussian Parameterization	w/o temporal axis (3×3)	0.1095	0.9435	9.7797	0.3036	0.8341	5.2206
	Diagonal precision (no coupling)	0.1094	0.9436	9.7841	0.2980	0.8375	5.3035
	Only spatial coupling	0.1096	0.9435	9.7791	0.2981	0.8374	5.3045
	Only spatio-temporal coupling	0.1095	0.9435	9.7810	0.2987	0.8371	5.2933
	Isotropic scalar σ	0.1205	0.9376	9.3355	0.3388	0.8128	4.7357
Brain Grid & Source Placement	Free-floating centers	0.1095	0.9435	9.7817	0.2976	0.8377	5.3094
	Surface-only grid	0.1370	0.9288	8.7703	0.6002	0.6316	2.2260
LD Conditioning	No conditioning (static)	0.9995	0.0112	0.0023	1.0000	0.0034	0.0001
	Global amplitude modulation	0.5414	0.6762	2.6841	0.2986	0.8371	5.2960
	Pre-interpolated LD input	0.1197	0.9381	9.3760	0.3014	0.8354	5.2520
Forward Model	Direct electrode mapping	0.2165	0.8845	6.9413	0.2831	0.8463	5.5331
	Learned linear projection	0.2122	0.8873	6.8383	0.5384	0.6791	2.6970
	Fixed leadfield matrix	0.2573	0.8614	5.9793	0.5955	0.6356	2.2600

5 Which channels should we keep? An SR-fidelity study of electrode subsets

Throughout the paper we have treated the low-density (LD) montage as a uniformly-random K-subset of the HD electrodes. In practice, sparse-EEG hardware imposes a *fixed* electrode layout, and a long line of EEG channel-selection work has argued that not all electrodes are equally informative for downstream tasks (Abdullah et al., 2022; Kim et al., 2022). We ask the *reverse* question that is natural for super-resolution: when the goal is to *reconstruct* the full montage from K retained channels, do these anatomically-motivated subsets remain a good choice, or does the inverse problem favor layouts with different geometry?

To that end, we define ten named subsets at three sparsification levels matching the SR factors used in the main benchmark (Visualization in figure 3, and full channel lists in Table 9, both in Appendix E): SR$\times$2 (K=31): Hemi-Left, Hemi-Right (left/right halves plus the four midline channels Fz, Cz, Pz, Oz); SR$\times$4 (K=15): V15 (rim-style superset of the ten circumference channels of Valderrama & Sheoran (2025)), FT15 (fronto-temporal), INT15 (central / midline control); SR$\times$8 (K=7): four hemispheric rim slices (VL7, VR7, VU7, VLw7) and an INT7 (central) control.

Results. Two findings stand out. *(i) Uniform spatial coverage dominates anatomical priors.* Across *every* tested (dataset, r) cell, uniform Random selection consistently achieves the lowest NMSE and highest PCC/SNR among the named subsets evaluated. The gap grows with the sparsification factor: at SR$\times$2 the named hemispheric subsets cost roughly $+0.16$–0.17 NMSE on both datasets, at SR$\times$4 roughly $+0.10$–0.14, and at SR$\times$8 roughly $+0.10$–0.15, and the seed-to-seed standard deviation of the named subsets (≤ 0.06 NMSE) is an order of magnitude smaller than the gap to Random(Table 3). We attribute this to the geometry of the inverse problem: a uniform random K-subset spreads observations isotropically over the scalp, while Anatomically-clustered subsets, by contrast, leave large portions of the scalp with no nearby observed electrode, and EMAG must extrapolate over those gaps.

(ii) Differences among named subsets are small and not monotonic in the rim-vs-interior axis. At SR$\times$4, the rim-style V15 and FT15 subsets slightly outperform the interior control INT15 on both datasets (NMSE gap ≈ 0.03). At SR$\times$8, however, the central INT7 is competitive or better than every rim slice; at this regime the absolute size K=7 is small enough that any clustered subset under-samples the rest of the head. The two SR$\times$2 hemispheric subsets are nearly indistinguishable (Table 3), in line with the small spatial sensitivity of the reconstructor at low SR factors.

6 Discussion and Conclusion

Why anisotropic 4D Gaussians work for EEG. Scalp EEG is a volume-conducted projection of cortical sources whose currents are oriented (perpendicular to the cortical sheet) and produce anisotropic scalp patterns that isotropic kernels cannot represent. The full 4×4 precision matrix

Table 3: EEG super-resolution fidelity on **SEED** and **SEED-IV** for each named electrode subset versus the uniform-random baseline. Metrics: NMSE (↓), PCC (↑), SNR in dB (↑). Per-cell Δ_{rand} numbers for all three metrics are reported in Table 10 of Appendix E.

		SEED (62 ch.)			**SEED-IV** (62 ch.)		
SR	Subset	NMSE ↓	PCC ↑	SNR (dB) ↑	NMSE ↓	PCC ↑	SNR (dB) ↑
×2	RANDOM	**0.154**±0.001	**0.919**±0.001	**8.20**±0.03	**0.115**±0.001	**0.940**±0.001	**9.58**±0.03
	HEMI-LEFT	0.312±0.058	0.829±0.035	5.15±0.77	0.285±0.030	0.845±0.018	5.50±0.45
	HEMI-RIGHT	0.312±0.050	0.829±0.031	5.13±0.68	0.275±0.032	0.851±0.019	5.67±0.50
×4	RANDOM	**0.278**±0.002	**0.850**±0.001	**5.62**±0.02	**0.203**±0.001	**0.893**±0.000	**7.06**±0.02
	V15	0.384±0.037	0.784±0.024	4.22±0.42	0.299±0.010	0.837±0.006	5.29±0.15
	FT15	0.385±0.038	0.783±0.024	4.19±0.42	0.311±0.020	0.829±0.012	5.12±0.28
	INT15	0.418±0.037	0.762±0.024	3.83±0.38	0.330±0.025	0.818±0.015	4.89±0.33
×8	RANDOM	**0.422**±0.003	**0.760**±0.002	**3.80**±0.03	**0.336**±0.001	**0.814**±0.001	**4.80**±0.01
	VL7	0.540±0.017	0.677±0.013	2.70±0.14	0.501±0.001	0.706±0.001	3.02±0.01
	VR7	0.527±0.001	0.687±0.001	2.80±0.01	0.491±0.000	0.713±0.000	3.10±0.00
	VU7	0.561±0.002	0.662±0.002	2.53±0.02	0.470±0.001	0.727±0.001	3.32±0.01
	VLw7	0.572±0.004	0.654±0.003	2.44±0.03	0.469±0.005	0.727±0.003	3.34±0.05
	INT7	0.529±0.006	0.685±0.005	2.78±0.05	0.441±0.001	0.747±0.001	3.61±0.01

captures this spatial anisotropy and couples it to the temporal evolution of source patterns in a single parameterization.

Comparison with prior reconstruction methods. Unlike MNE (Hämäläinen & Ilmoniemi, 1994) or sLORETA (Pascual-Marqui et al., 2002), EMAG learns its forward model jointly with the source parameters and needs no precomputed leadfield from structural MRI, so it applies even when individual anatomy is unavailable. Unlike electrode-space CNN/diffusion baselines (e.g. SRGDiff (Liu et al., 2025a)) it also exposes an explicit source configuration in 3D brain space, which we examine in App. H. Robustness to clinical-grade noise and artifacts is untested, as the present datasets are research-grade; a noise-stress evaluation is left to future work.

Model explainability. Because each component carries an explicit position $\mathbf{g}_n$ in 3D brain space, a precision matrix $\mathbf{\Sigma}_n^{-1}$, and a signed amplitude w_n, the trained model can be read out directly in source space. High-$|w_n|$ components cluster near the known stimulation sites in Localize-MI (App. H.1, Fig. 8); the full 4×4 covariance lets each component be classified as point-like, elongated (directed cortical current), or sheet-like (App. H.2, Fig. 9; top-K ellipsoids in Fig. 6); and the coupling terms c_3, c_4, c_5 let a Gaussian's spatial extent vary in time. This level of source-space readout is not available from CNN/diffusion SR baselines.

When uniform sampling beats anatomy. Channel-selection recipes from the classifier-driven EEG literature do not transfer to super-resolution: anatomically-clustered subsets leave large scalp regions unobserved and a generative reconstructor pays a measurable fidelity cost (Sec. 5). Layouts that maximize spatial coverage, which uniform random sampling approximates, are the better default for SR pipelines.

Limitations. EMAG is personalized by default; a single pooled/LOSO model nonetheless generalizes to unseen subjects at close to per-subject quality (Sec. 4.3). The spherical brain grid remains a simplification of the true brain geometry; future work could use subject-specific cortical meshes. Evaluation is currently focused on three datasets; broader validation on naturalistic EEG paradigms (e.g., resting state, BCI motor imagery) is warranted.

EMAG demonstrates that physics-inspired 4D Gaussian splatting offers a practical path for EEG super-resolution—lightweight personalized models that unlock HD insights from sparse clinical hardware. By bridging computer vision rendering techniques with EEG's ill-posed inverse problem, we hope EMAG opens new directions for neural field representations in neuroscience. All code and experiments will be made publicly available upon acceptance. Additionally, we provide an ethics and responsible-use discussion in App. K.

AI USE DISCLOSURE

In this work, we used generative AI tools to implement parts of the method and evaluation code, aggregate experimental results across seeds, and assist in interpreting results. We have not used generative AI tools to develop the method, formulate hypotheses, or design the experiments, and generating synthetic data, mathematical claims and proofs, translation, and qualitative data analysis are not applicable to this work. Additionally, we used generative AI tools to create or modify figures, draft and edit parts of the paper for readability and consistency, and format references. We have reviewed all AI-assisted work: all code was verified and tested by the authors, all reported numbers were cross-checked against the raw experiment outputs, and all text was reviewed and revised by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Abdullah, Ibrahima Faye, and Md Rafiqul Islam. Eeg channel selection techniques in motor imagery applications: A review and new perspectives. *Bioengineering*, 9(12):726, 2022.

Isaac A Corley and Yufei Huang. Deep eeg super-resolution: Upsampling eeg spatial resolution with generative adversarial networks. In *2018 IEEE EMBS international conference on biomedical & health informatics (BHI)*, pp. 100–103. IEEE, 2018.

Anders M Dale, Arthur K Liu, Bruce R Fischl, Randy L Buckner, John W Belliveau, Jeffrey D Lewine, and Eric Halgren. Dynamic statistical parametric mapping: combining fmri and meg for high-resolution imaging of cortical activity. *neuron*, 26(1):55–67, 2000.

Richard J Davidson. Anterior cerebral asymmetry and the nature of emotion. *Brain and cognition*, 20(1):125–151, 1992.

Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th international IEEE/EMBS conference on neural engineering (NER)*, pp. 81–84. IEEE, 2013.

Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d-rotor gaussian splatting: towards efficient novel view synthesis for dynamic scenes. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–11, 2024.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. Mne software for processing meg and eeg data. *neuroimage*, 86:446–460, 2014.

Roberta Grech, Tracey Cassar, Joseph Muscat, Kenneth P Camilleri, Simon G Fabri, Michalis Zervakis, Petros Xanthopoulos, Vangelis Sakkalis, and Bart Vanrumste. Review on solving the inverse problem in eeg source analysis. *Journal of neuroengineering and rehabilitation*, 5(1):25, 2008.

Hans Hallez, Bart Vanrumste, Roberta Grech, Joseph Muscat, Wim De Clercq, Anneleen Vergult, Yves D'Asseler, Kenneth P Camilleri, Simon G Fabri, Sabine Van Huffel, et al. Review on solving the forward problem in eeg source analysis. *Journal of neuroengineering and rehabilitation*, 4(1): 46, 2007.

Matti S Hämäläinen and Risto J Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & biological engineering & computing*, 32(1):35–42, 1994.

Gexin Huang, Ke Liu, Jiawen Liang, Chang Cai, Zheng Hui Gu, Feifei Qi, Yuanqing Li, Zhu Liang Yu, and Wei Wu. Electromagnetic source imaging via a data-synthesis-based convolutional encoder–decoder network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):6423–6437, 2022.

Mohammad Rafid Ul Islam, Prasad Tadepalli, and Alan Fern. Self-attention-based diffusion model for time-series imputation in partial blackout scenarios. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17564–17572, 2025.

Meng Jiao, Shihao Yang, Xiaochen Xian, Neel Fotedar, and Feng Liu. Multi-modal electrophysiological source imaging with attention neural networks based on deep fusion of eeg and meg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:2492–2502, 2024.

Shahar Ain Kedem, Itamar Zimerman, and Eliya Nachmani. Neural brain fields: A nerf-inspired approach for generating nonexistent eeg electrodes. *arXiv preprint arXiv:2601.00012*, 2025.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, et al. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

Minjae Kim, Seungjae Yoo, and Chul Kim. Miniaturization for wearable eeg systems: recording hardware and data processing. *Biomedical Engineering Letters*, 12(3):239–250, 2022.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Dongdong Li, Zhongliang Zeng, Zhe Wang, and Hai Yang. Estformer: Transformer utilizing spatiotemporal dependencies for eeg super-resolution. *arXiv e-prints*, pp. arXiv–2312, 2023.

Hongjun Liu, Leyu Zhou, Zijianghao Yang, and Chao Yao. Step-aware residual-guided diffusion for eeg spatial super-resolution. *arXiv preprint arXiv:2510.19166*, 2025a.

Renjie Liu, Zaijun Wang, Jiang Qiu, and Xue Wang. Assigning channel weights using an attention mechanism: an eeg interpolation algorithm. *Frontiers in neuroscience*, 17:1251677, 2023.

Zijin Liu, Xiang Zhao, and You Song. Rdpi: a refine diffusion probability generation method for spatiotemporal data imputation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12255–12263, 2025b.

Christoph M Michel and Denis Brunet. Eeg source imaging: a practical review of the analysis steps. *Frontiers in neurology*, 10:325, 2019.

Ezequiel Mikulan, Simone Russo, Sara Parmigiani, Simone Sarasso, Flavia Maria Zauli, Annalisa Rubino, Pietro Avanzini, Anna Cattani, Alberto Sorrentino, Steve Gibbs, et al. Simultaneous human intracerebral stimulation and hd-eeg, ground-truth for source localization methods. *Scientific data*, 7(1):127, 2020.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Roberto Domingo Pascual-Marqui et al. Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. *Methods find exp clin pharmacol*, 24(Suppl D):5–12, 2002.

Mario Quiles Pérez, Sergio López Bernal, Eduardo Horna Prat, Luis Montesano Del Campo, Lorenzo Fernández Maimó, and Alberto Huertas Celdrán. Eeg channel reconstruction using convolutional neural networks in limited bcis: A proposed method for neuromarketing applications. *Applied Soft Computing*, 181:113455, 2025.

François Perrin, Jacques Pernier, Olivier Bertrand, and Jean Francois Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology*, 72(2):184–187, 1989.

Rabie A Ramadan and Athanasios V Vasilakos. Brain computer interface: control signals review. *Neurocomputing*, 223:26–44, 2017.

Sari Saba-Sadiya, Tuka Alhanai, Taosheng Liu, and Mohammad M Ghassemi. Eeg channel interpolation using deep encoder-decoder networks. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2432–2439. IEEE, 2020.

Rui Sun, Abbas Sohrabpour, Gregory A Worrell, and Bin He. Deep neural networks constrained by neural mass models improve electrophysiological source imaging of spatiotemporal brain dynamics. *Proceedings of the National Academy of Sciences*, 119(31):e2201128119, 2022.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

Camilo E Valderrama and Anshul Sheoran. Identifying relevant eeg channels for subject-independent emotion recognition using attention network layers. *Frontiers in psychiatry*, 16:1494369, 2025.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Julius Vetter, Jakob H Macke, and Richard Gao. Generating realistic neurophysiological time series with denoising diffusion probabilistic models. *Patterns*, 5(9), 2024.

Shuqiang Wang, Tong Zhou, Yanyan Shen, Ye Li, Guoheng Huang, and Yong Hu. Generative ai enables eeg super-resolution via spatio-temporal adaptive diffusion learning. *IEEE Transactions on Consumer Electronics*, 71(1):1034–1045, 2025.

Jonathan R Wolpaw. Brain-computer interfaces (bcis) for communication and control. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pp. 1–2, 2007.

Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20310–20320, 2024.

Shunyang Zhang, Senzhang Wang, Xianzhen Tan, Ruochen Liu, Jian Zhang, and Jianxin Wang. Sasdim: Self-adaptive noise scaling diffusion model for spatial time series imputation. *arXiv preprint arXiv:2309.01988*, 2023.

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, 49(3): 1110–1122, 2018.

A Implementation details

This section details the architecture, initialization and training setup of Sec. 3.

Brain grid. The brain domain spans $[-90, 90]$ mm in each spatial dimension. The brain grid at resolution R consists of all points on a uniform $R \times R \times R$ lattice that fall within a sphere of radius 90 mm. For the default $R = 12$, the sphere grid yields $N \approx 672$ grid points. With $G = 3$ Gaussians per point, the default configuration has $N_{\text{total}} \approx 2{,}016$ components and $\sim$73k total learnable parameters.

Per-component parameterization. Each Gaussian has 12 learnable parameters: 1 base amplitude, 1 temporal center $\mu^{(t)}$, 4 Cholesky diagonal entries (log-scale), and 6 Cholesky off-diagonal entries. Total per-Gaussian parameters: $12 \times N_{\text{total}} \approx 24{,}192$ for the default configuration. The LD encoder adds $\sim$2K parameters; the modulation MLP (mapping $\mathbb{R}^{32} \to \mathbb{R}^{64} \to \mathbb{R}^{N}$) adds $\sim$43K parameters at $N \approx 672$, dominating the total parameter budget of $\sim$73k.

Computational efficiency. The Cholesky-based Mahalanobis decomposition (Eq. 4) enables EMAG to use the full 4D covariance at minimal overhead over isotropic variants. The spatial terms A and C

are precomputed once per forward pass, and only the scalar temporal displacement τ varies at each time step, making the per-timestep cost $\mathcal{O}(M \cdot N_{\text{total}})$.

Initialization. Base amplitudes w_n are initialized to zero. Temporal centers $\mu_n^{(t)}$ are drawn from $\text{Uniform}(0, 0.5)$ (normalized time). Cholesky diagonal entries are initialized to $\ell_i = \log(1/15)$ for spatial dimensions (a spatial standard deviation of 15 mm, matching the grid spacing $180/12$ mm so neighbouring components overlap) and $\ell_3 = \log(1/0.1)$ for the temporal dimension. Off-diagonal Cholesky entries are initialized to zero (no coupling at initialization).

LD encoder. The temporal encoder is a 1D convolution over the LD channels at each time step, producing a $d_f = 32$-dimensional feature vector. The modulation MLP has one hidden layer with 64 units and ReLU activation, mapping from $\mathbb{R}^{d_f}$ to $\mathbb{R}^N$ (one value per grid point, shared across G Gaussians).

Training protocol. All models are trained for up to 100 epochs with early stopping (patience 20, minimum improvement 10^{-4}) using Adam (Kingma & Ba, 2014) (learning rate 10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) with gradient clipping at $\|\nabla\|_2 \leq 1$. Each epoch is processed in temporal chunks with gradient accumulation. Training uses a single NVIDIA RTX 6000 Ada (48GB) GPU and completes in 30–120 minutes per subject for the default configuration.

Compute resources. All experiments ran on a single internal SLURM cluster (NVIDIA RTX6000 GPUs with ∼48 GB VRAM, single-GPU allocation per job; 4 CPU cores, 24 GB RAM). A single (subject, SR-factor) EMAG training run takes 30–120 min wall-clock at r=2–16. The full benchmark (3 datasets × 3–4 SR factors × 5 seeds × 7–15 subjects) plus the ablation grid (Section F) consumed approximately 180 GPU-hours; preliminary and failed runs (hyperparameter search, debugging) add an estimated additional $1-2\times$.

B DATASETS

This section details the preprocessing, splits and licenses of the datasets in Sec. 4.1.

Dataset preprocessing and split protocol. For Localize-MI we use the released averaged-reference 256-ch HD-EEG at 8 kHz, retain the ± 200 ms window around stimulation onset, and apply a per-subject z-score with statistics computed on the training split only. For SEED and SEED-IV we use the preprocessed 0–75 Hz band-pass signals at 200 Hz released with each dataset, segment into non-overlapping 4 s windows (800 samples), and apply per-channel z-scoring with train-split statistics. The 80/20 train/test split is by trial within subject; the random split seed is fixed per run and we report mean±std across five seeds $s \in \{0, 1, 2, 3, 4\}$ unless otherwise noted. The LD electrode subset is sampled uniformly without replacement from the HD montage, once per subject (RNG seed 0) and held fixed across all five split seeds and across all sessions of a subject. Windows are non-overlapping, so no sample is shared between train and test. Following standard convention, all metrics are computed over the full HD electrode set (both observed and reconstructed), matching the baselines.

Datasets, licenses and access. Localize-MI (Mikulan et al., 2020): released under CC BY 4.0; available from OpenNeuro (`ds003644`). **SEED** (Zheng & Lu, 2015) and **SEED-IV** (Zheng et al., 2018): released under their original BCMI laboratory data-use agreement; available upon application from bcmi.sjtu.edu.cn/home/seed. **Baselines.** SaSDim, SADI, RDPI, DDPMEEG, ESTformer, STAD and SRGDiff are used under their respective original licenses; we cite the source papers in Section 4.2 and Section 2 and release no modified weights.

C DOWNSTREAM-TASK UTILITY

To test whether reconstruction fidelity carries over to a downstream task, we train the same SEED three-class emotion classifier on two inputs: the sparse LD signal and EMAG's HD reconstruction from that same LD signal. Both use the identical per-subject split. EMAG's reconstruction improves accuracy over the LD input at every super-resolution factor: $+0.201$ ($2\times$), $+0.207$ ($4\times$), and $+0.169$ ($8\times$). The gain is positive for all $15/15$ subjects (paired t-test, $p < 10^{-8}$). It stays large even at $8\times$, where only 7 electrodes are observed, so the spatial structure EMAG recovers is task-relevant rather than cosmetic. The super-resolution step therefore gives a measurable and uniform task-level gain.

D CROSS-SUBJECT GENERALIZATION: EXTENDED RESULTS

This section gives the full protocol, per-(dataset, r) numbers with standard deviations, and additional discussion accompanying Section 4.3.

D.1 SINGLE-TO-SINGLE TRANSFER

Protocol. For every dataset $\mathcal{D} \in \{\text{SEED}, \text{SEED-IV}\}$ and SR factor $r \in \{2, 4, 8\}$ we use the per-subject EMAG checkpoints already trained for the main results in Table 1 (architecture: R=12, G=3, spherical brain grid, random LD selection, 5 split seeds). Let S denote the number of subjects in $\mathcal{D}$ (S=15 for both SEED and SEED-IV). Localize-MI is omitted here because its electrode positions are digitized per subject rather than sharing a standard cap; it is covered by the pooled and LOSO models (App. D.2). For each split seed s we form the $S \times S$ transfer matrix $T^{(s)}$ where $T^{(s)}_{ij}$ is obtained by:

1. loading the EMAG model trained on subject i (split seed s);
2. feeding it subject j's LD test recordings (the same LD subset and test split that was used when evaluating subject j in the main results, so $T^{(s)}_{jj}$ reproduces the within-subject number for that seed);
3. recording NMSE, PCC and SNR between the model's output and subject j's HD ground truth.

We report

$$\textit{within}^{(s)} \;=\; \tfrac{1}{S}\sum_{i} T^{(s)}_{ii}, \qquad \textit{cross}^{(s)} \;=\; \tfrac{1}{S(S-1)}\sum_{i\neq j} T^{(s)}_{ij},$$

and average across seeds. Each non-diagonal cell $T^{(s)}_{ij}$ uses no information from subject j during training: subject i's 4D Gaussian field is applied verbatim to subject j's LD signal, with subject j's electrode positions used only to evaluate the rendered output (the LD-conditioning MLP is, by design, shared across grid points but not across subjects).

LD-spline reference. As a model-free baseline we use spherical-spline interpolation: each LD sample (on the K=n_{hd}/r retained electrodes) is upsampled to the full HD montage by fitting a Legendre-polynomial spherical spline (m=4, λ=10^{-5}) per time step. The resulting HD prediction depends only on subject j's own LD signal and electrode positions, so it is by construction a cross-subject method (no training subject involved); we therefore report a single LD-spline column rather than a transfer matrix.

Full results. Table 4 reports mean±std across 5 split seeds for each (dataset, r). The variance is small (typically $< 10^{-2}$ on PCC), so all conclusions in Section 4.3 hold under any single seed.

Cross-subject transfer matrices Figure 2 shows the full 15×15 subject×subject PCC transfer matrices used to derive the within/cross/LD-spline summary in Table 4. The bright diagonal vs. pale off-diagonal pattern is consistent across both datasets and all SR factors, showing that a single-subject fit is largely subject-specific; multi-subject training (App. D.2) closes most of this gap.

Interpretation. Three observations are worth emphasizing. *(i) Cross-subject EMAG strictly dominates LD-spline on all tested (dataset, r, metric) configurations.* The improvement is largest on SEED (PCC $0.46-0.54$ vs. $0.26-0.28$; NMSE $\sim0.78-0.86$ vs. $\sim1.44-1.49$) where the LD-spline baseline is weak, and remains clear on SEED-IV. This says that the physics-style structure baked into EMAG—a sphere-bounded mixture of anisotropic 4D Gaussians whose centers act as virtual cortical sources, projected to scalp through subject electrode geometry—generalizes across people without any retraining. *(ii) The within-vs-cross gap is large* (e.g. on SEED-IV r=2, PCC $0.94 \rightarrow 0.57$, SNR $9.6 \rightarrow 1.1$ dB). Per-subject training thus captures, beyond the geometric prior, individual features of the source configuration (cortical folding, electrode-to-cortex distances, time-locked dipolar dynamics) that cannot be reused across subjects. In the per-subject fit the component centers are subject-specific; a pooled/LOSO model instead shares them across subjects and, as shown below, still reconstructs unseen subjects at close to per-subject quality. *(iii) The cross-subject score is nearly flat in r* (SEED-IV PCC $0.565\rightarrow0.549\rightarrow0.487$; SEED $0.539\rightarrow0.535\rightarrow0.460$), whereas the within-subject score degrades sharply with r. This is consistent with cross-subject performance being limited by the structural mismatch between two brains rather than by the LD information content—the latter only becomes the bottleneck once subject identity is matched.

Table 4: Full cross-subject results (mean±std over 5 split seeds). *Within* is the diagonal of the $S \times S$ transfer matrix (same protocol as Table 1). *Cross* is the off-diagonal mean ($S(S-1)$ ordered pairs per seed). *LD-spline* is non-learned spherical-spline upsampling of the LD montage. NMSE↓, PCC↑, SNR↑ (dB).

Dataset	r	Metric	EMAG within	EMAG cross	LD-spline
SEED	2×	NMSE	$0.1545_{\pm 0.0011}$	$0.8203_{\pm 0.0023}$	$1.4895_{\pm 0.0037}$
		PCC	$0.9194_{\pm 0.0006}$	$0.5392_{\pm 0.0013}$	$0.2552_{\pm 0.0019}$
		SNR	$8.20_{\pm 0.03}$	$0.92_{\pm 0.01}$	$-1.72_{\pm 0.01}$
	4×	NMSE	$0.2776_{\pm 0.0015}$	$0.7819_{\pm 0.0020}$	$1.4447_{\pm 0.0039}$
		PCC	$0.8496_{\pm 0.0009}$	$0.5349_{\pm 0.0011}$	$0.2776_{\pm 0.0019}$
		SNR	$5.62_{\pm 0.02}$	$1.13_{\pm 0.01}$	$-1.58_{\pm 0.01}$
	8×	NMSE	$0.4027_{\pm 0.0032}$	$0.8613_{\pm 0.0017}$	$1.4371_{\pm 0.0032}$
		PCC	$0.7595_{\pm 0.0021}$	$0.4603_{\pm 0.0014}$	$0.2815_{\pm 0.0016}$
		SNR	$3.80_{\pm 0.03}$	$0.72_{\pm 0.01}$	$-1.56_{\pm 0.01}$
SEED-IV	2×	NMSE	$0.1151_{\pm 0.0011}$	$0.7848_{\pm 0.0042}$	$1.0625_{\pm 0.0022}$
		PCC	$0.9405_{\pm 0.0006}$	$0.5651_{\pm 0.0016}$	$0.4687_{\pm 0.0011}$
		SNR	$9.59_{\pm 0.04}$	$1.13_{\pm 0.02}$	$-0.24_{\pm 0.01}$
	4×	NMSE	$0.2023_{\pm 0.0008}$	$0.7840_{\pm 0.0025}$	$1.1350_{\pm 0.0015}$
		PCC	$0.8928_{\pm 0.0004}$	$0.5493_{\pm 0.0004}$	$0.4325_{\pm 0.0007}$
		SNR	$7.07_{\pm 0.02}$	$1.13_{\pm 0.01}$	$-0.52_{\pm 0.01}$
	8×	NMSE	$0.3358_{\pm 0.0011}$	$0.8519_{\pm 0.0011}$	$1.2668_{\pm 0.0011}$
		PCC	$0.8143_{\pm 0.0007}$	$0.4870_{\pm 0.0007}$	$0.3666_{\pm 0.0005}$
		SNR	$4.80_{\pm 0.01}$	$0.79_{\pm 0.01}$	$-0.94_{\pm 0.01}$

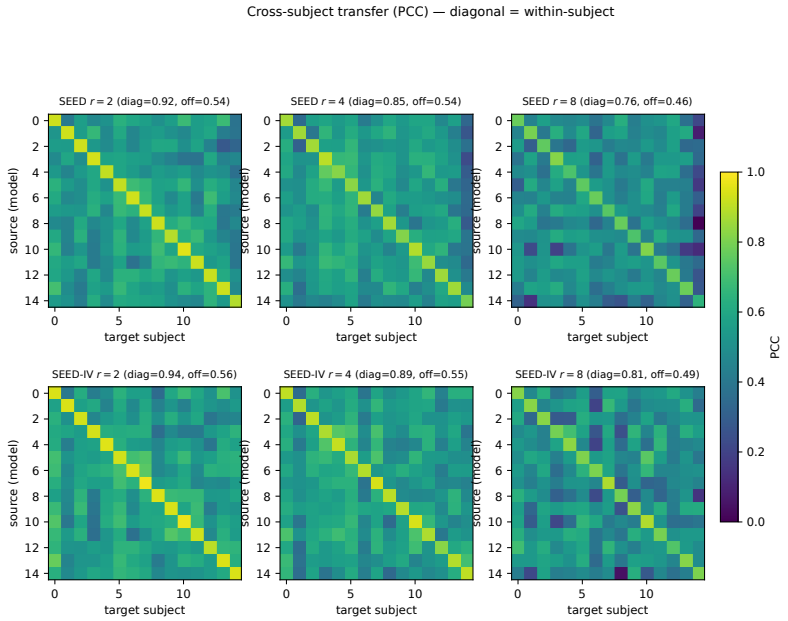

Figure 2: Subject×subject PCC transfer matrices. Rows: source (training) subject; columns: target (evaluation) subject. Diagonal = within-subject score; off-diagonal mean = cross-subject score reported in Table 4.

Table 5: Multi-subject generalization: per-subject baseline vs. pooled and leave-one-subject-out (LOSO) NMSE ↓. Mean±std over seeds (LOSO also over held-out subjects).

Dataset	r	Per-subject	Pooled	LOSO
Localize-MI	2	0.087	$0.101_{\pm 0.001}$	$0.115_{\pm 0.030}$
Localize-MI	4	0.105	$0.121_{\pm 0.001}$	$0.137_{\pm 0.033}$
Localize-MI	8	0.140	$0.153_{\pm 0.001}$	$0.172_{\pm 0.041}$
Localize-MI	16	0.213	$0.213_{\pm 0.001}$	$0.237_{\pm 0.051}$
SEED	2	0.156	$0.157_{\pm 0.001}$	$0.160_{\pm 0.037}$
SEED	4	0.280	$0.307_{\pm 0.002}$	$0.312_{\pm 0.051}$
SEED	8	0.424	$0.438_{\pm 0.003}$	$0.440_{\pm 0.061}$
SEED-IV	2	0.119	$0.123_{\pm 0.001}$	$0.129_{\pm 0.024}$
SEED-IV	4	0.208	$0.250_{\pm 0.002}$	$0.269_{\pm 0.047}$
SEED-IV	8	0.340	$0.360_{\pm 0.001}$	$0.374_{\pm 0.055}$

D.2 Pooled and leave-one-subject-out models

The single-to-single transfer analyzed earlier is a deliberately hard probe (one subject's model applied to one other). We additionally train genuine multi-subject models: a *pooled* model fit jointly on all subjects' training splits, and a *leave-one-subject-out* (LOSO) model trained on every subject but one and evaluated on the held-out subject. Table 5 reports NMSE against the per-subject baseline.

Three trends emerge. First, the unseen-subject penalty is small and grows gently with the SR factor: at $r{=}2$ the pooled/LOSO models are essentially at per-subject quality (SEED 0.157 vs. 0.156; Localize-MI 0.101/0.115 vs. 0.087), widening only modestly by $r{=}8$. Second, pooling helps over LOSO exactly where a subject is unseen: the pooled model (which has seen the evaluated subject during training) is consistently at or below the LOSO model, and on SEED/SEED-IV pooled stays within ≈ 0.04 NMSE of the per-subject fit. Third, on Localize-MI both multi-subject models remain below the prior SOTA SRGDiff (0.238 at $r{=}4$) at every factor, so even a completely unseen subject is reconstructed better than the best prior method trained conventionally. Per-subject fitting stays the highest-fidelity default, but the geometric prior clearly transfers, so one model can serve multiple subjects including new ones.

D.3 Per-subject reconstruction quality

Tables 6 – 8 report per-subject NMSE/PCC/SNR for every SR factor, using the default $R{=}12$, $G{=}3$ sphere-grid configuration at split seed 0. Per-subject variability is moderate on Localize-MI (e.g. at $r{=}4$, NMSE ranges from 0.07 for sub-02 to 0.18 for sub-04, the latter being a known outlier with denser implant coverage in motor cortex), and tighter on SEED/SEED-IV where the shared 62-channel cap normalizes input geometry. Reconstruction quality degrades smoothly with r: on Localize-MI the mean PCC drops $0.956 \to 0.946 \to 0.927 \to 0.885$ from $r{=}2$ to $r{=}16$.

Table 6: Per-subject EMAG reconstruction on **Localize-MI** (split seed 0, $R{=}12$, $G{=}3$, sphere grid). NMSE ↓, PCC ↑, SNR ↑ (dB).

Subject	2×			4×			8×			16×		
	NMSE	PCC	SNR	NMSE	PCC	SNR	NMSE	PCC	SNR	NMSE	PCC	SNR
1	0.068	0.965	11.67	0.086	0.956	10.64	0.110	0.943	9.59	0.161	0.916	7.93
2	0.053	0.973	12.75	0.070	0.964	11.53	0.107	0.945	9.69	0.179	0.906	7.47
3	0.070	0.964	11.52	0.086	0.956	10.65	0.126	0.935	8.99	0.177	0.907	7.52
4	0.154	0.920	8.13	0.182	0.905	7.41	0.222	0.882	6.54	0.316	0.827	5.00
5	0.066	0.966	11.80	0.081	0.959	10.93	0.114	0.941	9.44	0.190	0.900	7.20
6	0.097	0.950	10.12	0.114	0.941	9.43	0.150	0.922	8.24	0.253	0.864	5.97
7	0.091	0.953	10.39	0.108	0.945	9.68	0.156	0.919	8.08	0.229	0.878	6.40
Mean	**0.086**	**0.956**	**10.91**	**0.104**	**0.946**	**10.04**	**0.141**	**0.927**	**8.65**	**0.215**	**0.885**	**6.78**

Table 7: Per-subject EMAG reconstruction on **SEED** (split seed 0, R=12, G=3, sphere grid). NMSE ↓, PCC ↑, SNR ↑ (dB).

Subject	2×			4×			8×		
	NMSE	PCC	SNR	NMSE	PCC	SNR	NMSE	PCC	SNR
1	0.148	0.923	8.29	0.256	0.863	5.92	0.403	0.773	3.95
2	0.142	0.926	8.49	0.262	0.859	5.81	0.365	0.797	4.38
3	0.143	0.926	8.43	0.257	0.862	5.91	0.428	0.756	3.69
4	0.121	0.937	9.16	0.193	0.898	7.14	0.333	0.817	4.77
5	0.173	0.910	7.63	0.332	0.817	4.79	0.485	0.718	3.14
6	0.177	0.907	7.53	0.319	0.825	4.96	0.434	0.752	3.63
7	0.136	0.929	8.65	0.262	0.859	5.82	0.405	0.771	3.92
8	0.163	0.915	7.88	0.300	0.837	5.23	0.453	0.740	3.44
9	0.144	0.925	8.41	0.260	0.860	5.85	0.410	0.768	3.88
10	0.209	0.889	6.80	0.337	0.814	4.72	0.483	0.719	3.16
11	0.088	0.955	10.55	0.214	0.887	6.70	0.345	0.810	4.63
12	0.157	0.918	8.05	0.285	0.846	5.46	0.445	0.745	3.52
13	0.164	0.914	7.86	0.274	0.852	5.62	0.406	0.771	3.92
14	0.149	0.923	8.27	0.267	0.856	5.74	0.396	0.777	4.02
15	0.220	0.883	6.58	0.375	0.790	4.26	0.580	0.648	2.36
Mean	**0.156**	**0.919**	**8.17**	**0.280**	**0.848**	**5.59**	**0.425**	**0.757**	**3.76**

Table 8: Per-subject EMAG reconstruction on **SEED-IV** (split seed 0, R=12, G=3, sphere grid). NMSE ↓, PCC ↑, SNR ↑ (dB).

Subject	2×			4×			8×		
	NMSE	PCC	SNR	NMSE	PCC	SNR	NMSE	PCC	SNR
1	0.110	0.943	9.57	0.183	0.904	7.36	0.333	0.817	4.78
2	0.111	0.943	9.56	0.217	0.885	6.64	0.324	0.822	4.90
3	0.133	0.931	8.77	0.192	0.899	7.17	0.323	0.823	4.91
4	0.136	0.930	8.67	0.231	0.877	6.37	0.373	0.792	4.28
5	0.096	0.951	10.16	0.168	0.912	7.74	0.314	0.828	5.03
6	0.130	0.933	8.85	0.241	0.871	6.18	0.352	0.805	4.53
7	0.113	0.942	9.45	0.216	0.885	6.66	0.353	0.804	4.52
8	0.048	0.976	13.15	0.127	0.934	8.95	0.228	0.879	6.42
9	0.103	0.947	9.85	0.195	0.897	7.10	0.374	0.791	4.27
10	0.120	0.938	9.21	0.208	0.890	6.81	0.319	0.825	4.96
11	0.061	0.969	12.15	0.117	0.940	9.32	0.225	0.880	6.47
12	0.138	0.928	8.60	0.231	0.877	6.36	0.366	0.796	4.36
13	0.196	0.897	7.07	0.324	0.822	4.89	0.470	0.728	3.28
14	0.107	0.945	9.71	0.217	0.885	6.64	0.328	0.820	4.84
15	0.112	0.942	9.50	0.182	0.905	7.41	0.357	0.802	4.47
Mean	**0.114**	**0.941**	**9.62**	**0.203**	**0.892**	**7.04**	**0.336**	**0.814**	**4.80**

E ELECTRODE-SUBSET SUPER-RESOLUTION: EXTENDED RESULTS

This section backs the electrode-subset study of Section 5. We give (i) the prior-work motivation behind each named subset, (ii) the full channel lists, (iii) the experimental protocol, and (iv) per-(dataset, SR-factor, subset) fidelity numbers including SNR.

E.1 ELECTRODE GEOMETRY

Figure 3 shows the three recording geometries to scale: SEED and SEED-IV share a 62-channel scalp cap, while Localize-MI uses a 256-channel scalp HD-EEG cap with subject-specific digitized positions (sub-01 shown). This visual contrast underlies many of the dataset-dependent patterns in the ablation study (Sec. 4.4), including the larger benefit of the volumetric brain grid on SEED.

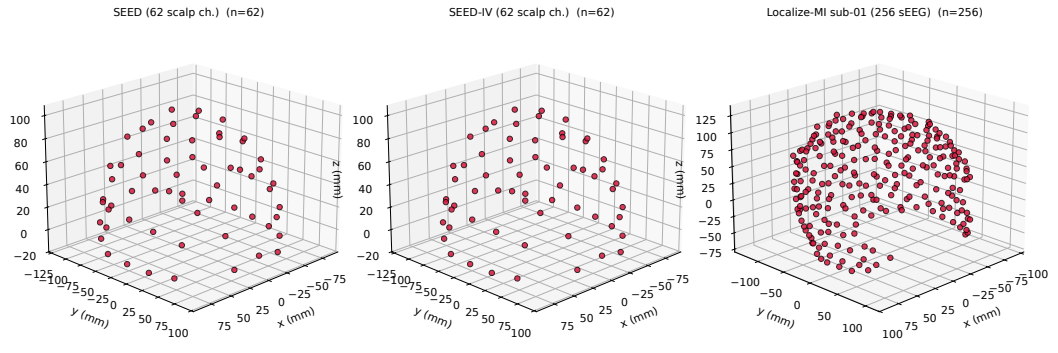

Figure 3: Recording geometries used in this paper. SEED/SEED-IV (left, middle) use the standard 10/20-extended 62-channel scalp cap; Localize-MI sub-01 (right) uses a 256-channel scalp HD-EEG montage.

E.2 BACKGROUND AND MOTIVATION FOR THE NAMED SUBSETS

Channel selection has been an active subfield of EEG-based BCI design for over two decades; recent surveys review supervised, filter-, and wrapper-based methods that reduce the number of required channels without sacrificing classifier accuracy (Abdullah et al., 2022; Kim et al., 2022). For affective EEG specifically, two complementary priors recur in the literature.

Frontal hemispheric asymmetry. Davidson (1992) established the now-canonical link between asymmetric frontal alpha activity and approach / withdrawal motivation, which has motivated emotion-EEG datasets such as SEED (Zheng & Lu, 2015) to emphasize frontal and lateral coverage. Our HEMI-LEFT / HEMI-RIGHT subsets at $K{=}31$ operationalize this prior by retaining one full hemisphere of the montage plus the four midline channels (Fz, Cz, Pz, Oz).

Circumference / rim channels. Valderrama & Sheoran (2025) argue that ten "circumference" channels of the 10/05 montage (Fp1, Fp2, F7, F8, T7, T8, P7, P8, O1, O2) carry most of the information used by direct emotion classifiers, and report that classifiers trained on these ten channels match the accuracy of classifiers trained on the full 62-channel montage on SEED. We use this finding as the inspiration for our V15 subset (the ten circumference channels plus five additional rim electrodes, yielding $K{=}15$) and for the four hemispheric rim slices at $K{=}7$ (VL7, VR7, VU7, VLw7). The FT15 subset is a fronto-temporal variant that emphasizes channels typically associated with affective processing.

Interior controls. For each rim subset we include a same-K interior control (INT15, INT7) consisting of central / midline electrodes. These controls let us isolate the effect of *rim versus interior* placement at fixed budget.

Subset definitions Each named subset is a list of channel labels from the canonical 62-channel SEED / SEED-IV ordering. Subset sizes are $K \in \{31, 15, 7\}$, giving SR factors of $\{2, 4, 8\}$ relative to the full montage. Full description of channels in Table 9 and a top-down view of the montages in Figure 4.

Table 9: Definitions of the named electrode subsets. V15 is a superset of the ten circumference channels of Valderrama & Sheoran (2025). HEMI-LEFT and HEMI-RIGHT retain the four midline channels (Fz, Cz, Pz, Oz) in both subsets to keep the size at exactly $K=31$.

Subset ID	K	Channels (SEED order)
HEMI-LEFT	31	FP1, AF3, F7, F5, F3, F1, FT7, FC5, FC3, FC1, T7, C5, C3, C1, TP7, CP5, CP3, CP1, P7, P5, P3, P1, PO7, PO5, PO3, CB1, O1, FZ, CZ, PZ, OZ
HEMI-RIGHT	31	FP2, AF4, F8, F6, F4, F2, FT8, FC6, FC4, FC2, T8, C6, C4, C2, TP8, CP6, CP4, CP2, P8, P6, P4, P2, PO8, PO6, PO4, CB2, O2, FZ, CZ, PZ, OZ
V15	15	FP1, FP2, F7, F8, FT7, FT8, T7, T8, TP7, TP8, P7, P8, PO7, O1, O2
FT15	15	FP1, FP2, F3, F4, F7, F8, FC5, FC6, FT7, FT8, T7, T8, TP7, TP8, P7
INT15	15	AF3, AF4, F1, F2, FZ, FC1, FC2, C1, C2, CZ, CP1, CP2, P1, P2, PZ
VL7	7	FP1, F7, FT7, T7, TP7, P7, O1
VR7	7	FP2, F8, FT8, T8, TP8, P8, O2
VU7	7	FP1, FP2, F7, F8, FT7, FT8, T7
VLw7	7	TP7, TP8, P7, P8, PO7, O1, O2
INT7	7	FZ, FC1, FC2, CZ, CP1, CP2, PZ

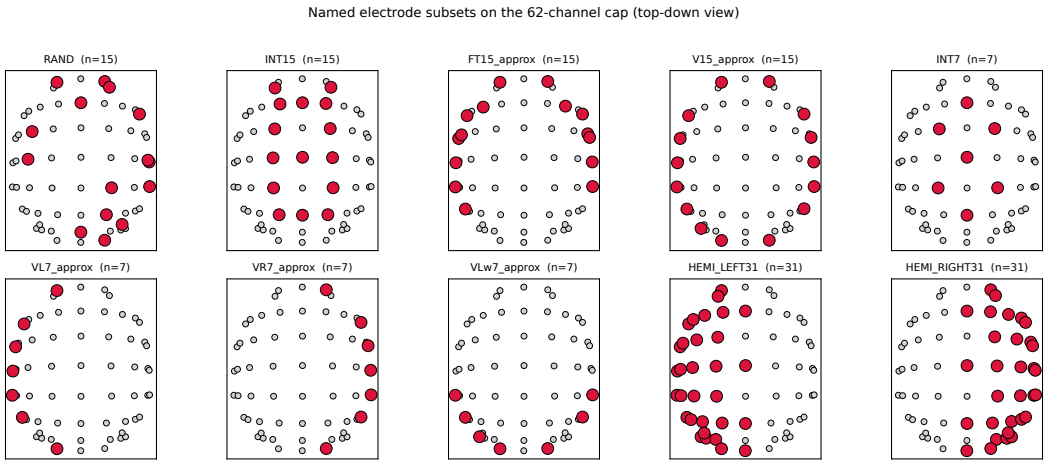

Figure 4: Named electrode subsets on the 62-channel SEED/SEED-IV cap (top-down view, anterior up). Crimson = retained channels, gray = discarded.

E.3 PROTOCOL

For every (dataset, SR-factor, subset) cell we trained EMAG with the standard configuration of the main paper: mixture of $G{=}3$ anisotropic 4D Gaussians per active grid point, 12^3 uniform brain grid, full-4D covariance, differentiable Gaussian forward model, batch size 1, 100 epochs with patience 20 on validation NMSE. The *only* quantity that varies across cells is the set of retained low-density channels: for the named subsets, this set is fixed to the channel list of Table 9; for the RANDOM baseline, the K-subset is sampled uniformly at random per (subject, split seed). Every cell is trained from scratch – the RANDOM baseline does *not* share weights with the main-benchmark random-subset model – so the Random-versus-named comparison is a clean test of channel selection isolated from any compute-budget confound.

We repeat each (dataset, SR-factor, subset, subject) configuration across three split seeds $s \in \{0, 1, 2\}$, which control both the train/val/test split and – for the RANDOM baseline – the random K-subset. The numbers reported in Section 5 and in Table 10 below are the mean over the 15 subjects of the dataset, with mean$\pm$std taken across the three split seeds.

E.4 FULL RECONSTRUCTION-FIDELITY TABLE

Table 10 reports NMSE, PCC, and SNR for every (dataset, SR, subset) cell as mean$\pm$std over the three split seeds. The Δ_{rand} columns are absolute changes relative to the RANDOM baseline at the same (dataset, SR) cell; positive Δ_{rand}NMSE and negative Δ_{rand}PCC / Δ_{rand}SNR all indicate that the named subset is worse than RANDOM.

Table 10: Three-seed reconstruction-fidelity numbers for the electrode-subset experiments. Each cell is the mean over the 15 subjects of the dataset, reported as mean$\pm$std across split seeds $s \in \{0, 1, 2\}$. The RANDOM rows are the same uniform-random baseline employed throughout the paper, retrained at each split seed with an independently re-sampled K-subset per (subject, seed).

			Metric (mean$\pm$std over 3 seeds)			Δ_{rand} (vs. RANDOM, same SR)		
Dataset	SR	Subset	NMSE ↓	PCC ↑	SNR (dB) ↑	ΔNMSE	ΔPCC	ΔSNR
SEED	×2	RANDOM	0.154±0.001	0.919±0.001	8.20±0.03	0.000	0.000	0.00
SEED	×2	HEMI-LEFT	0.312±0.058	0.829±0.035	5.15±0.77	+0.158	−0.090	−3.05
SEED	×2	HEMI-RIGHT	0.312±0.050	0.829±0.031	5.13±0.68	+0.158	−0.091	−3.08
SEED	×4	RANDOM	0.278±0.002	0.850±0.001	5.62±0.02	0.000	0.000	0.00
SEED	×4	V15	0.384±0.037	0.784±0.024	4.22±0.42	+0.106	−0.066	−1.41
SEED	×4	FT15	0.385±0.038	0.783±0.024	4.19±0.42	+0.108	−0.066	−1.44
SEED	×4	INT15	0.418±0.037	0.762±0.024	3.83±0.38	+0.140	−0.087	−1.79
SEED	×8	RANDOM	0.422±0.003	0.760±0.002	3.80±0.03	0.000	0.000	0.00
SEED	×8	INT7	0.529±0.006	0.685±0.005	2.78±0.05	+0.108	−0.074	−1.01
SEED	×8	VR7	0.527±0.001	0.687±0.001	2.80±0.01	+0.105	−0.072	−0.99
SEED	×8	VL7	0.540±0.017	0.677±0.013	2.70±0.14	+0.118	−0.082	−1.09
SEED	×8	VU7	0.561±0.002	0.662±0.002	2.53±0.02	+0.139	−0.098	−1.26
SEED	×8	VLW7	0.572±0.004	0.654±0.003	2.44±0.03	+0.150	−0.106	−1.35
SEED-IV	×2	RANDOM	0.115±0.001	0.940±0.001	9.58±0.03	0.000	0.000	0.00
SEED-IV	×2	HEMI-RIGHT	0.275±0.032	0.851±0.019	5.67±0.50	+0.160	−0.090	−3.91
SEED-IV	×2	HEMI-LEFT	0.285±0.030	0.845±0.018	5.50±0.45	+0.170	−0.095	−4.08
SEED-IV	×4	RANDOM	0.203±0.001	0.893±0.000	7.06±0.02	0.000	0.000	0.00
SEED-IV	×4	V15	0.299±0.010	0.837±0.006	5.29±0.15	+0.097	−0.056	−1.77
SEED-IV	×4	FT15	0.311±0.020	0.829±0.012	5.12±0.28	+0.108	−0.063	−1.94
SEED-IV	×4	INT15	0.330±0.025	0.818±0.015	4.89±0.33	+0.127	−0.075	−2.17
SEED-IV	×8	RANDOM	0.336±0.001	0.814±0.001	4.80±0.01	0.000	0.000	0.00
SEED-IV	×8	INT7	0.441±0.001	0.747±0.001	3.61±0.01	+0.105	−0.067	−1.19
SEED-IV	×8	VLW7	0.469±0.005	0.727±0.003	3.34±0.05	+0.134	−0.087	−1.46
SEED-IV	×8	VU7	0.470±0.001	0.727±0.001	3.32±0.01	+0.135	−0.087	−1.48
SEED-IV	×8	VR7	0.491±0.000	0.713±0.000	3.10±0.00	+0.156	−0.102	−1.70
SEED-IV	×8	VL7	0.501±0.001	0.706±0.001	3.02±0.01	+0.165	−0.108	−1.78

F ABLATIONS: EXTENDED RESULTS AND DISCUSSION

This section gives the full ablation results and per-component discussion underlying Section 4.4. We conduct a systematic ablation study to validate each core design decision in EMAG. All ablations are evaluated on the Localize-MI and SEED datasets using the $4\times$ super-resolution factor, as this setting provides a representative balance between reconstruction difficulty and spatial ambiguity. We report NMSE, PCC, and SNR on the held-out test set, averaged across all subjects. Results are summarized in the Table 2.

F.1 HYPERPARAMETER SEARCH

We searched grid resolution $R \in \{5, 7, 10, 12, 15\}$ and Gaussians-per-point $G \in \{1, 2, 3, 5, 8\}$ on Localize-MI sub-01 at r=4 with a TPE search (OPTUNA); learning rate was jointly sampled in $\{1\times10^{-4}, \ldots, 5\times10^{-3}\}$. For every (R, G) pair we report the lowest validation NMSE achieved across the sampled learning rates.

Table 11: Hyperparameter search on Localize-MI (r=4): validation NMSE (↓) by grid resolution R (rows) and Gaussians-per-point G (columns). Bold = chosen configuration.

$R\backslash G$	1	2	3	5	8
5	0.150	0.164	0.244	0.405	0.405
7	0.142	0.143	0.132	0.274	0.136
10	0.131	0.131	0.115	0.128	0.136
12	0.123	0.114	**0.106**	0.131	0.115
15	0.113	0.113	0.113	0.131	0.107

Performance saturates by R=12, with diminishing returns at R=15 (and $\sim 2\times$ compute). Increasing G beyond 3 does not improve fidelity at $R \geq 10$. Learning-rate sensitivity was narrow: all top-quartile trials used $\mathrm{lr} \in [5\times10^{-4}, 3\times10^{-3}]$; we adopt $\mathrm{lr}=1\times10^{-3}$ throughout. Final headline configuration: R=12, G=3 (**bold**).

Because EMAG is a personalized, per-subject model, we further asked whether the best configuration is itself subject-specific. First, ranking each grid point's components by amplitude, the lowest-amplitude one still carries $\approx$15.5% of the mixture mass in every one of the 30 SEED/SEED-IV subjects, so no component collapses to zero amplitude and the mixture is used the same way across subjects. Second, an independent per-subject hyperparameter search (Optuna, r=4,Localize-MI subjects) returns optima that cluster in a single region — $R \in \{15, 16\}$ (mean 15.8), $G \in \{5, \ldots, 8\}$ (mean 6.0), learning rate 5×10^{-4}–1.2×10^{-3}, mean validation NMSE 0.108 — rather than scattering across subjects, and improve on the global R=12, G=3 model by only a small NMSE margin at several times its parameters and training time. Both point the same way: the component budget need not be made subject-specific, and a single shared configuration is appropriate.

F.2 4D GAUSSIAN PARAMETERIZATION

The distinguishing structural element of EMAG is its use of full 4×4 anisotropic precision matrices for each Gaussian component. We isolate the contribution of each aspect of this parameterization through three targeted ablations.

Temporal dimension in the precision matrix. We first ask whether the fourth (temporal) axis of the precision matrix carries information beyond what amplitude modulation alone provides. To test this, we replace the full 4×4 precision matrix with a 3×3 spatial-only block, removing the temporal dimension from the Mahalanobis distance computation in Eq. 2 entirely. In this variant, temporal dynamics are handled exclusively by the time-varying amplitude $a_n(t)$.

We find that removing the temporal axis consistently degrades performance across both datasets. The larger relative degradation on SEED is consistent with the richer spectral content of continuous emotion-elicitation recordings, where transient activations with complex spatiotemporal profiles are more common than in the short, stereotyped evoked responses of Localize-MI. In both cases the degradation confirms that the temporal covariance axis encodes information that amplitude modulation alone cannot recover.

Off-diagonal coupling terms. We next isolate the contribution of anisotropy and space-time coupling by zeroing out all off-diagonal entries $c_0, \ldots, c_5$ in the Cholesky factor $\mathbf{L}_n$ (Eq. 3), reducing each component to an axis-aligned ellipsoid. We further decompose this into two sub-ablations: (a) removing only the spatial off-diagonal terms c_0, c_1, c_2, preserving spatial isotropy but retaining space-time coupling; and (b) removing space-time terms c_3, c_4, c_5 while introducing spatial anisotropy.

Results across both datasets reveal a consistent pattern: removing *any* subset of off-diagonal terms produces similar degradation, and the two sub-ablations produce comparable lesser results. Two conclusions follow. First, both coupling types are *jointly* necessary: either type alone recovers no advantage over the fully uncoupled case, indicating that their contributions are complementary rather than additive. Second, the full 4×4 precision matrix, which enables all six coupling terms simultaneously, is strictly required to capture the interaction between source directionality and temporal evolution, and no lower-dimensional subset achieves an equivalent representation.

Isotropic scalar variance. As the minimal parameterization within the Gaussian family, we replace each component's full precision matrix with a single learnable scalar width σ_n, yielding a fully isotropic Gaussian with one variance parameter per component. This corresponds to the isotropic variant described in Section 4.2 and establishes the performance floor for the Gaussian family.

The isotropic variant produces the largest degradation within the parameterization category. Taken together with the off-diagonal ablations, the results establish a clear ordering: full $4 \times 4 >$ diagonal anisotropic $>$ isotropic, with each step yielding a consistent performance improvement. The steeper decline on SEED suggests that the anisotropic parameterization is particularly important when source configurations are spatially distributed and heterogeneous, as in emotion-elicitation EEG, compared to the more focal, stereotyped activations of the stimulation paradigm in Localize-MI.

F.3 Brain Grid and Source Placement

EMAG anchors its Gaussian components on an anatomically-constrained spherical brain grid. We evaluate the necessity of this structural prior and its performance with other grid variations.

Anatomical grid constraint. We remove the spherical constraint of Eq. 1 and allow all Gaussian centers to be free-floating learnable parameters, initialized at the original grid positions but unconstrained during training. Without the anatomical prior, both reconstruction quality and training stability degrade. Inspecting learned positions, a subset of free-floating components migrates outside the brain volume entirely, attaching to the scalp surface. These surface-hugging components act as shortcut interpolators, exploiting proximity to LD electrodes rather than learning meaningful source configurations. This phenomenon confirms that the brain grid constraint is not merely a computational convenience but a necessary regularizer that prevents the model from collapsing to a degenerate surface interpolation solution.

Cortical surface vs. volumetric grid. To test whether deep (subcortical) sources contribute to reconstruction accuracy, we restrict grid points to the brain surface by retaining only those grid points within a thin shell of the outer sphere boundary. This surface-only model closely approximates cortical surface source models used in classical EEG analysis.

The impact of restricting sources to the brain surface is markedly dataset-dependent and considerably stronger than anticipated. On Localize-MI, performance degrades measurably, which is expected with the dataset's known subcortical stimulation sites that the surface-only model cannot represent. On SEED, however, the degradation is even greater, a finding which challenges the standard assumption that emotion-related EEG activity is predominantly cortical: the volumetric subcortical grid appears to be critical for effective SEED reconstruction. This ablation establishes that the full volumetric grid is the appropriate default and that surface-constrained source models are not a viable alternative for general-purpose EEG super-resolution.

F.4 Low-Density Conditioning Architecture

The LD conditioning pipeline is responsible for translating sparse observed electrode signals into spatially-resolved source amplitude modulations. We verify each component of this pipeline.

No conditioning (static source template). We first establish the most critical baseline: removing the TemporalEncoder and MLP_{mod} entirely, leaving each Gaussian with only its fixed base amplitude

$w_{i,g}$ and no dependence on the LD input. In this degenerate setting, the model becomes a fixed anatomical template with learned spatial geometry but no ability to respond to the observed signal.

As expected, performance collapses under this model. This result confirms the self-evident but important fact that input-dependent conditioning is the primary driver of per-sample reconstruction accuracy, and that the Gaussian source geometry alone, while beneficial for generalization, cannot substitute for dynamic signal-driven modulation. This ablation also serves as an interpretable lower bound: any method that approaches this performance level is effectively failing to use the LD input.

Global vs. spatially-resolved amplitude modulation. We replace the per-grid-point amplitude output $\Delta w_i(t) \in \mathbb{R}^N$ (Eq. 8) with a single global scalar $\Delta w(t) \in \mathbb{R}$ applied uniformly to all grid points. This preserves the temporal conditioning mechanism but collapses spatial resolution: the model knows *when* the brain is active but not *where*.

The contrast between datasets here is the most striking in the entire ablation suite. On Localize-MI, collapsing the per-grid-point modulation to a global scalar causes a near-catastrophic degradation, while on SEED the same modification has a lesser effect. This dissociation directly reflects the nature of the two tasks. Localize-MI recordings consist of focal, spatially-specific evoked responses to single-site intracranial stimulation. SEED recordings, however, capture distributed, slower-varying emotion-related activity across the entire cortex. This result highlights an important design consideration: the benefit of spatially-resolved conditioning scales with the spatial heterogeneity and focal specificity of the neural signals, and is most critical in clinical or evoked-response settings.

Raw vs. pre-interpolated LD input. We replace the raw m-channel LD input to the TemporalEncoder with a spherical-spline-interpolated version upsampled to the full M HD electrode positions (Perrin et al., 1989). This tests whether the encoder is bottlenecked by the sparsity of its input, and whether classical interpolation as a pre-processing step could substitute for the implicit spatial regularization provided by the Gaussian forward model.

We found that pre-interpolation yields no improvement over raw LD input. We attribute this result to the Gaussian source model acting as a strong spatial prior: because the forward model constrains electrode signals to arise from smooth volumetric source distributions, the encoder receives implicit spatial regularization through the backward pass during training, making external pre-interpolation redundant. This finding has a practical implication: EMAG does not require any electrode-space pre-processing and can operate directly on raw LD recordings.

F.5 Physics-Inspired Forward Model

The differentiable Gaussian rendering pipeline is the architectural element that most distinguishes EMAG from prior electrode-space super-resolution methods. We evaluate its necessity and examine alternatives.

Direct electrode-space mapping. We replace the entire source field rendering pipeline with a direct mapping: the TemporalEncoder output $\mathbf{h}_t$ is passed to an MLP that directly predicts the M-dimensional HD signal at each timestep, with no intermediate source representation. Model capacity is approximately matched to EMAG by widening the MLP hidden dimension. This ablation represents the strongest possible "physics-free" baseline within our framework and directly tests the central premise of the paper.

The benefit of the physics-inspired Gaussian pipeline is dataset-specific but always present. On Localize-MI, the direct mapping degrades with a substantial gap that confirms the value of the source-space bottleneck for focal, evoked-response data. On SEED, the gap narrows, which is consistent with the relatively smooth and distributed nature of emotion-related EEG: when the target signal lacks sharp spatial gradients, a capacity-matched MLP could learn an adequate electrode-space mapping from data alone. The contrast across datasets underscores that the physics-inspired design is most valuable precisely in the settings where it is clinically important: when neural activity is spatially focal and the super-resolution fidelity is critical.

Learned linear projection vs. Gaussian kernel. We replace the Gaussian-weighted distance function in Eq. 2 with a learned linear matrix $\mathbf{W} \in \mathbb{R}^{M \times N_{\text{total}}}$, trained end-to-end alongside the rest of the model. This preserves the source-space bottleneck representation but replaces the physically-

grounded Gaussian kernel with an unconstrained linear projection, testing whether the specific *form* of the Gaussian kernel is necessary or whether any differentiable source-to-electrode mapping suffices.

The learned linear projection behaves very differently across the two datasets, revealing an important generalization asymmetry. On Localize-MI, its scores nearly match those from the direct electrode mapping, landing far below EMAG. This indicates that the source-space bottleneck alone does not confer a meaningful advantage when the projection kernel is unconstrained. On SEED, the learned projection performs dramatically worse, and we hypothesize this stems from over-parameterization: the $M \times N_{\text{total}}$ weight matrix has no geometric prior and is free to memorize subject-specific electrode relationships. With 15 subjects and longer continuous recordings, the SEED training set exposes a wider variety of spatial configurations. Thus, the unconstrained projection fails to generalize across this diversity, whereas EMAG's Gaussian kernel imposes a distance-decay inductive bias that is invariant to specific electrode layouts and subjects. Together, the two datasets demonstrate that the *form* of the Gaussian kernel, and not merely the existence of a source-space intermediate, greatly contributes to EMAG's generalization capabilities.

Classical leadfield forward model. Finally, we replace the learned Gaussian kernel with a fixed electromagnetic leadfield matrix $\mathbf{L} \in \mathbb{R}^{M \times N_{\text{total}}}$ derived from a standard three-shell spherical head model using MNE-Python (Gramfort et al., 2014), keeping all other components of EMAG unchanged. The leadfield is computed at the same grid point locations as EMAG's brain grid, providing a direct comparison between physics-derived and learned projection kernels.

The fixed leadfield model represents the lowest-performing forward model variant on both datasets. We argue that the performance gap relative to EMAG arises from multiple compounding factors: (1) the spherical three-shell head model does not account for individual anatomical variation, (2) the fixed kernel cannot adapt to the data, and (3) the BEM-derived conductivity assumptions do not hold for all subjects or electrode montages. Notably, however, the leadfield model does not improve interpretability in the manner one might expect from a physically-derived kernel, since the reconstruction error itself becomes large enough to obscure meaningful source configurations. This result reinforces the core design choice of EMAG: a *learned* Gaussian kernel that is geometrically structured but data-adaptive outperforms both the fully unconstrained and the fully fixed alternatives, occupying a productive middle ground between flexibility and physical inductive bias.

F.6 Ablation Summary

Table 2 consolidates all ablation results. Three overarching findings emerge. First, the single most impactful component is spatially-resolved LD conditioning: removing it entirely drives NMSE to ≈ 1.0 on both datasets, while collapsing it to a global scalar causes a $+434\%$ NMSE increase on Localize-MI, underscoring that dynamic, spatially-specific modulation is indispensable for focal neural activity. Second, the full 4×4 precision matrix contributes meaningfully beyond isotropic baselines: the complete parameterization reduces NMSE by 0.0192 on Localize-MI and 0.0612 on SEED relative to a scalar-width Gaussian, with both spatial anisotropy and space-time coupling terms jointly required. Third, the physics-inspired Gaussian kernel's advantage over the direct mapping support our claim that embedding the EEG forward model as a differentiable rendering step is particularly beneficial for super-resolution, where data-driven interpolation is often constrained. Taken together, these results validate each major design decision in EMAG: anatomical grid anchoring, volumetric source placement, full 4D anisotropic covariance, and spatially-resolved conditioning are all necessary components, each contributing independently to the overall performance.

G Data efficiency

This section reports how reconstruction fidelity scales with per-subject training data. We probe how much per-subject data EMAG needs by shrinking the training set while holding the test set fixed. We keep the standard per-subject $80/20$ trial split and always evaluate on the same 20% test trials; a *train fraction* f then selects a fraction f of all trials (drawn from the fixed 80% training pool) for fitting, so $f{=}0.8$ is the full standard split and $f{=}0.2$ uses a quarter of the available training pool. Since the evaluation trials never change, differences reflect data size alone. Reconstruction is remarkably stable as data shrinks: on Localize-MI at $r{=}2$, NMSE rises only from 0.087 (full) to 0.090 at $f{=}0.2$, 0.095 at $f{=}0.1$, and 0.103 at $f{=}0.05$ (Fig. 5) — a sixteen-fold data reduction costs under 0.02 NMSE.

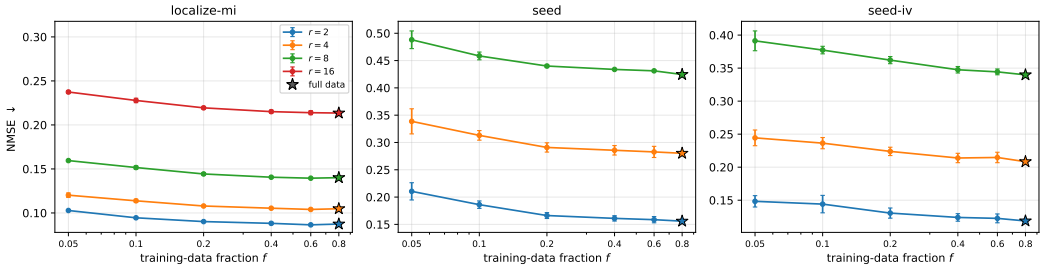

Figure 5: Data efficiency: NMSE vs. training-data fraction ($\star$ = full-data per-subject baseline); Mean over 5 seeds.

Combined with the model's small size ($\sim$73k parameters per subject, under two hours on a single GPU), EMAG is thus efficient in both parameters and data. Wall-clock time to convergence scales roughly linearly with the inverse of the training fraction: halving the training data approximately doubles the time to early stopping, because smaller training sets need more epochs to converge.

H Exploring the learned Gaussian sources

This section analyzes the learned Gaussian sources referenced in Sec. 6. The learned amplitudes of the $\sim 2{,}000$ Gaussian components reveal where the model concentrates explanatory power inside the brain volume. Figure 8 renders the active grid points of two trained subjects (one per dataset) colored by mean amplitude magnitude across the $G{=}3$ Gaussians per point. On SEED (sub-01), high-amplitude components form a distributed pattern with clear bilateral occipital and frontal foci, matching the prior on emotion-related EEG. On Localize-MI (sub-01), the learned amplitudes concentrate near the intracerebral implant sites, consistent with the focal stimulation paradigm of the dataset.

H.1 Learned Gaussian source maps

Quantitative validation against stimulation sites. We test whether the trained mixture allocates capacity near the ground-truth source. For each Localize-MI subject we load the trained checkpoint ($r{=}4$, default config) and score every Gaussian by its LD-conditioned post-stim energy

$$s_n \;=\; \mathbb{E}_{\text{trial},\, t\in[\,t_{\text{stim}},\,t_{\text{end}}\,]}\big[(w_n + m_n(\mathrm{LD}(t)))^2\big],$$

where w_n is the base amplitude, $m_n(\cdot)$ the LD modulator output (Section 3.4), and t_{stim} the stimulation onset of the run. For every run we then compute the Euclidean distance (mm, T1w-native space) from each top-K ($K{=}10$) Gaussian center to the stimulation site, defined as the midpoint $(\mathbf{p}_{e_1} + \mathbf{p}_{e_2})/2$ of the two adjacent intracranial contacts used for bipolar stimulation. As a chance baseline we draw $5{,}000$ random points from the brain grid and report the mean distance and an empirical *percentile* $\Pr[d_{\text{rand}} \le d_{\text{top1}}]$ (fraction of random brain-grid points lying closer to the stimulation site than the top Gaussian); this is a percentile, not a significance test. Per-run distances are averaged within subject; we then report the mean and standard deviation across the seven subject means.

Across 7 subjects and 61 stim runs (Table 12), the nearest top-Gaussian sits at $d_1 = 42.1 \pm 14.4$ mm from the stim midpoint, vs. a chance distance of 93.3 ± 5.5 mm; the top-1 Gaussian lands in the best $10 \pm 8\%$ of the brain volume on average. The effect is consistent across subjects (per-subject d_1: 22–64 mm) and is achieved *without any source-localization supervision* during training. We stress 42 mm is not precise localization — only concentration markedly closer than chance — and the stimulation site is not the sole source: intracranial stimulation drives network-level propagation, so scalp signals reflect multiple regions. A robustness check that ranks Gaussians by $|w_n|$ alone (no LD modulation) gives a near-identical headline ($d_1 = 42.3 \pm 11.9$ mm, $p = 0.09 \pm 0.06$), confirming the finding does not depend on the conditioning term. Figure 6 visualizes the top-10 Gaussian ellipsoids against the stim midpoints, and Figure 7 shows the corresponding reconstructed scalp-amplitude topographies in the post-stim window with the stim-midpoint projection (black $\star$) and the top-10 Gaussian projections (lime $\bullet$, size $\propto$ score) overlaid.

Table 12: Distance (mm) from learned top-K Gaussians to ground-truth sEEG stimulation midpoints, Localize-MI (r=4, default config), K=10, ranked by post-stim LD-conditioned energy. d_1: nearest top-Gaussian; $\bar{d}_K$: mean over the top-K; d_{chance}: mean distance from a random brain-grid point; p: empirical fraction of random points closer than d_1. Overall row = mean±std across the seven subject means.

Subject	d_1 (mm) ↓	$\bar{d}_K$ (mm)	d_{chance} (mm)	p
sub-01	**22.0**	53.0	93.5	**0.02**
sub-02	44.1	66.8	101.8	0.06
sub-03	64.5	83.1	90.1	0.23
sub-04	54.9	97.1	89.9	0.18
sub-05	38.7	72.1	89.4	0.08
sub-06	29.4	64.0	88.2	0.04
sub-07	40.9	56.5	100.2	0.09
Overall	**42.1±14.4**	**70.4±15.4**	**93.3±5.5**	**0.10±0.08**

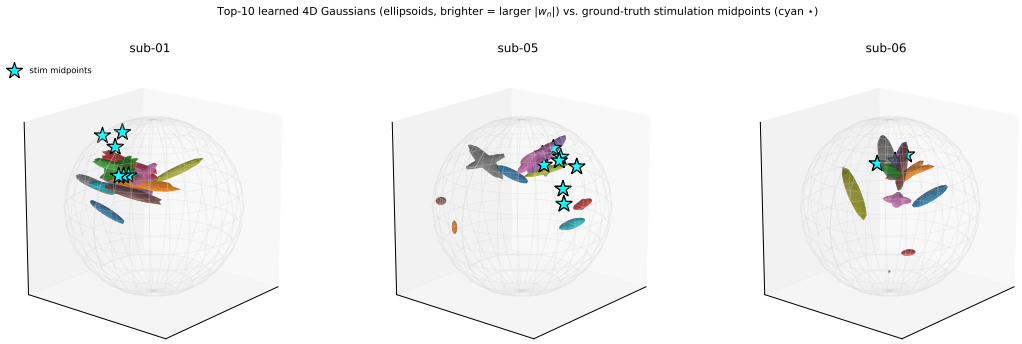

Figure 6: Top-10 learned 4D Gaussians rendered as 1σ spatial ellipsoids (brightness $\propto |w_n|$) versus ground-truth stimulation midpoints (cyan $\star$), in T1w-native space inside the brain envelope (r=90 mm). Subjects shown: best (sub-01), median (sub-05), and worst (sub-06).

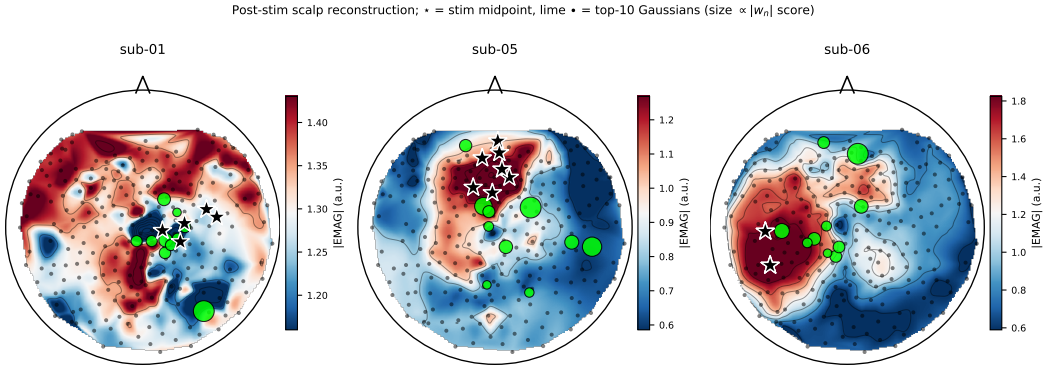

Figure 7: Reconstructed scalp amplitude in the post-stim window (mean |EMAG| over 0.25–0.26 s, across test trials). Black $\star$ = scalp electrode nearest each run's stim midpoint; lime $\bullet$ = top-10 Gaussians projected azimuthally to the same layout, marker size $\propto$ post-stim LD-conditioned score. Reconstructed amplitude peaks consistently sit near both the stim-midpoint projection and the high-scoring Gaussian projections.

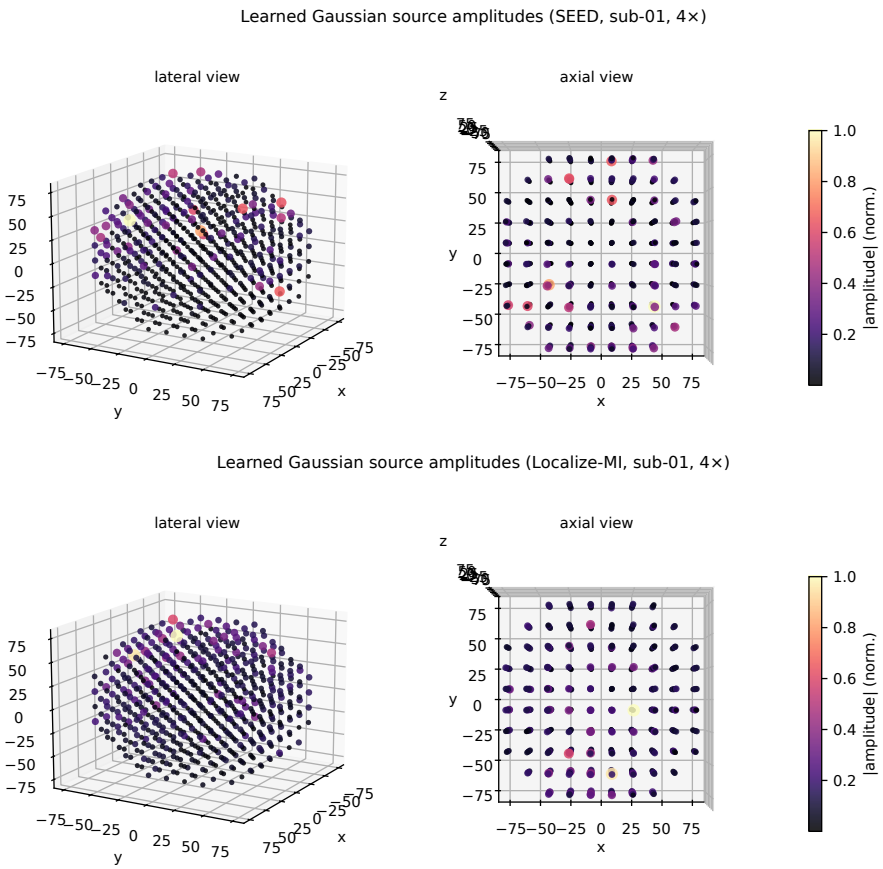

Figure 8: Learned Gaussian amplitudes (lateral and axial views) for SEED sub-01 (top) and Localize-MI sub-01 (bottom), trained at $r{=}4$ with the default configuration. Color and size encode the mean amplitude magnitude across the three Gaussians anchored at each grid point.

H.2 ANISOTROPY DISTRIBUTION

Figure 9 shows distributions of (a) the log-condition-number of the spatial 3×3 covariance block, (b) the per-axis spatial standard deviation, and (c) the temporal standard deviation of the 4×4 Cholesky-parameterized covariance. The anisotropy histogram is heavy-tailed and clearly separated from 0 (isotropic), confirming that components do not collapse to spheres at training time and supporting the parameterization ablation in Sec. F.2. Spatial spreads concentrate around $5-15$ mm, in line with the 15 mm initialization, while temporal spreads cover the full normalized range, indicating that the temporal axis of the precision matrix specializes per component—a behavior the diagonal/spatial-only ablations cannot reproduce.

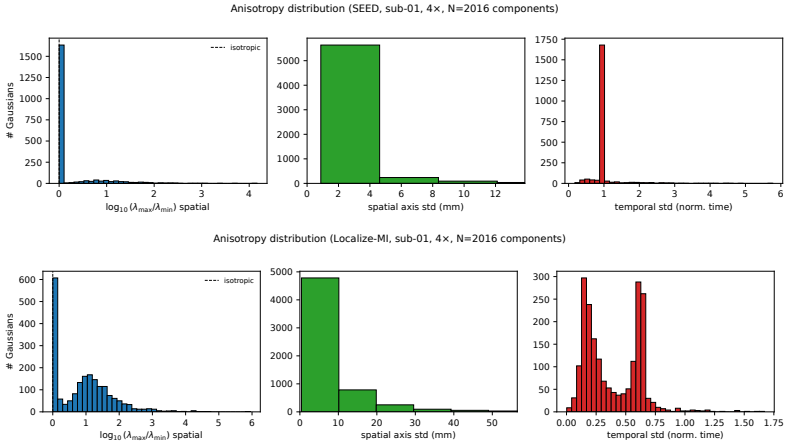

Figure 9: Anisotropy of the trained 4D Gaussian field (top: SEED sub-01, bottom: Localize-MI sub-01, both at $r{=}4$). Left: spatial log-condition-number $\log_{10}(\lambda_{\max}/\lambda_{\min})$ of the 3×3 spatial covariance block. Middle: per-axis spatial standard deviation in millimetres. Right: temporal standard deviation in normalized time units.

I QUALITATIVE EXAMPLES

This section shows qualitative reconstructions supporting Sec. 4.2.

I.1 WAVEFORM RECONSTRUCTION EXAMPLES

Figure 10 plots a representative high-density channel of sub-01 at every SR factor. EMAG tracks the slow envelope and the millisecond-scale stimulation artifacts at $r{=}2-8$, with visible amplitude attenuation at $r{=}16$ where only 16 input channels remain.

I.2 SPECTRAL FIDELITY

Figure 11 compares Welch power-spectral densities of the true HD signal and EMAG's reconstruction, averaged over channels and trials. EMAG matches the broadband shape and reproduces the dataset-specific peaks (Localize-MI stimulation harmonics; SEED α band around 10 Hz), with mild high-frequency attenuation that is consistent with the smoothness prior of the Gaussian forward model.

I.3 SINGLE-TIME-STEP SNAPSHOT

Figure 12 visualizes one HD time step as a top-down electrode scatter, comparing the sparse LD input fed to EMAG, the EMAG HD reconstruction, and the ground-truth HD field. The reconstruction recovers spatial structure that is invisible in the LD input alone, including sign reversals across hemispheres on SEED.

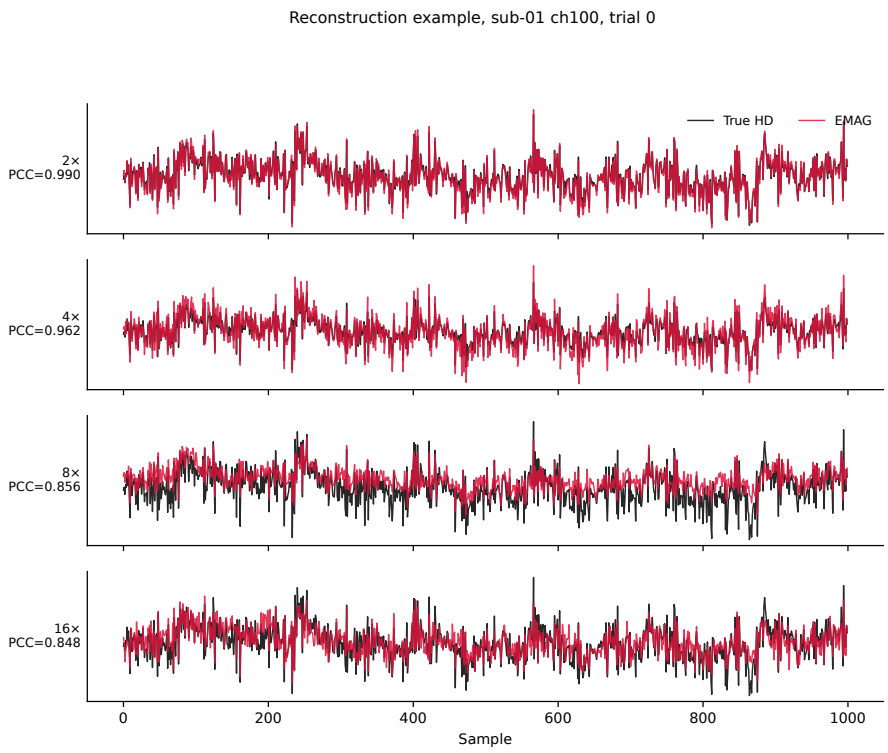

Figure 10: Localize-MI sub-01, channel 100, trial 0: ground-truth high-density signal (black) vs. EMAG reconstruction (crimson) at every SR factor. Per-trace PCC reported on the left.

J Limitations and Impact

J.1 Limitations

EMAG is personalized by default; a single pooled or LOSO model transfers to unseen subjects at close to per-subject quality on the same electrode montage (Section 4.3), but transfer across different montages is untested. Evaluation is restricted to non-pathological scalp EEG; behavior on clinical recordings, MEG, or intracranial signals is uncharacterized. Source-localization fidelity is externally validated only on Localize-MI, where ground-truth stimulation sites are available. Compute scales with the number of active grid points; sub-second inference on edge hardware would require further optimization.

J.2 Broader impact

EMAG can lower the equipment cost of high-fidelity scalp EEG by recovering HD-quality recordings from sparse, consumer-grade caps, with positive implications for low-resource clinical and BCI use. The same capability introduces risks: reconstructions are model-based extrapolations, not direct measurements, and may create spurious confidence in clinical decisions taken on low-density recordings. We recommend downstream-task validation (e.g. classification accuracy, ERP-component fidelity) before any clinical deployment, and we explicitly discourage use of EMAG outputs as standalone diagnostic evidence.

K Ethics and Responsible Use

This work uses only publicly released EEG datasets that were originally collected under the ethical approvals and licenses reported by their creators. We did not collect any new human-subject data, interact with participants, or perform any additional interventions beyond the use of these existing

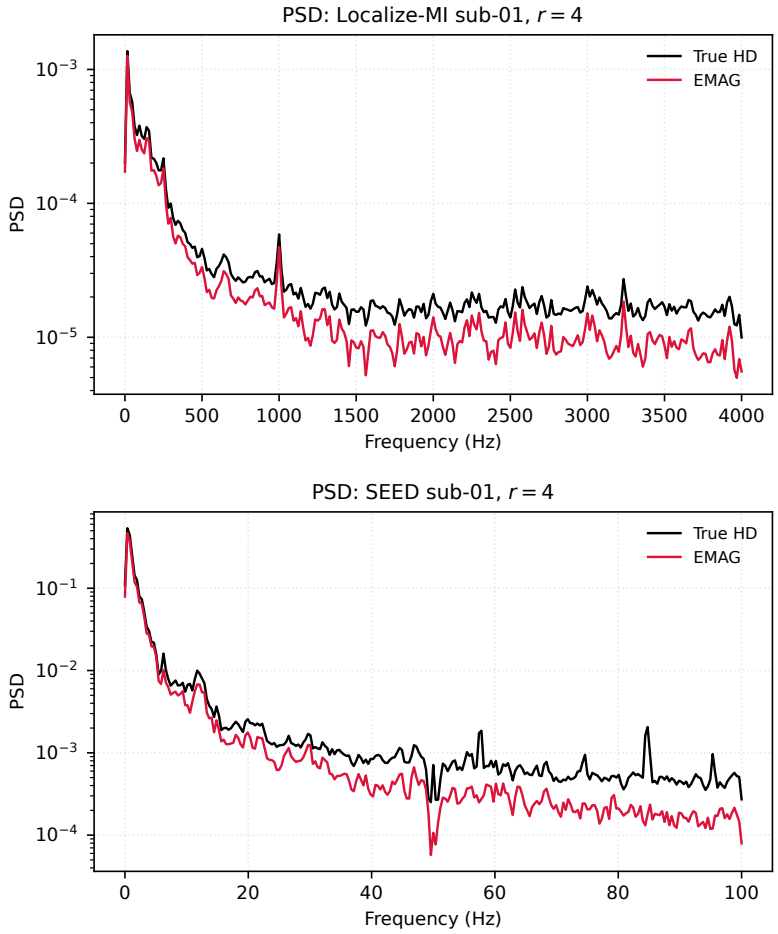

Figure 11: Welch PSD: true (black) vs. EMAG (crimson). Top: Localize-MI sub-01 (f_s=8 kHz); bottom: SEED sub-01 (f_s=200 Hz). Both at r=4.

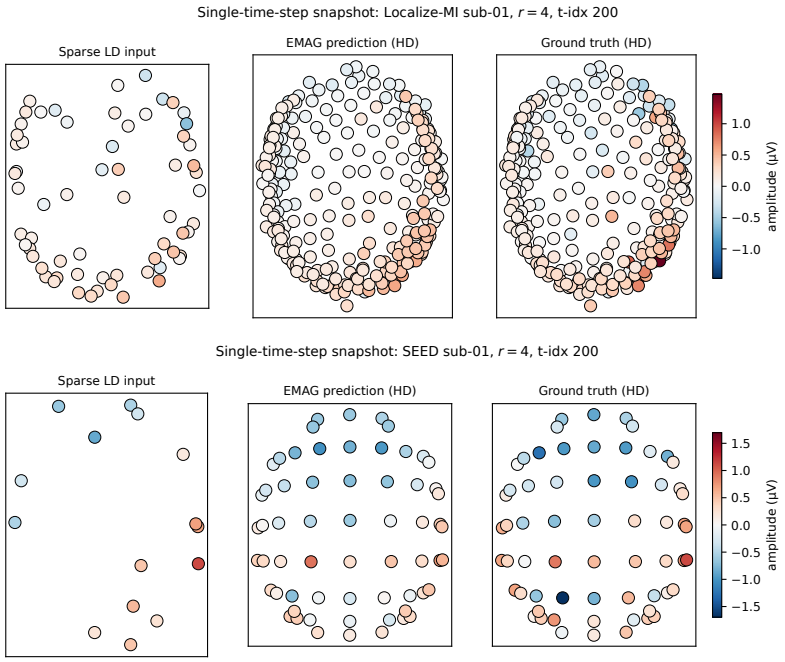

Figure 12: Single-time-step electrode-space snapshots (top: Localize-MI sub-01, bottom: SEED sub-01, r=4). Left: sparse LD input ($K=\lfloor n_{\mathrm{hd}}/r \rfloor$ retained channels). Middle: EMAG HD reconstruction. Right: ground truth.

de-identified datasets. Because EEG signals are biometric and can potentially be sensitive, we treat reconstruction outputs as research artifacts rather than diagnostic evidence, and we caution that they should not be used for clinical decision-making without proper validation, regulatory review, and domain-specific oversight. More broadly, high-fidelity EEG super-resolution may improve low-cost neurotechnology and retrospective analysis, but it could also be misused to over-interpret sparse recordings or create unwarranted confidence in reconstructed signals. For this reason, we recommend that any downstream deployment clearly disclose the reconstruction process, report uncertainty where possible, and validate conclusions on the original observed data rather than on reconstructed signals alone.