MART: STRUCTURED REGIONAL ADAPTATION FOR EEG REPRESENTATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Adapting EEG representation models to downstream tasks is often treated as a parameter-efficient update problem over the representation interface exposed by the backbone. We argue that, for EEG, the interface itself can be a limiting factor because flat hidden states may obscure the channel-time provenance and montage-dependent measurement conditions from which EEG representations are derived. We propose MART, a Measurement-Aware Regional Transformation framework that transforms this exposed interface into a structured regional adaptation space before lightweight downstream adaptation. Given EEG representations with channel-time provenance, MART recovers their channel-time organization, reparameterizes them into scalp-region representations, and augments each region with coverage and availability descriptors. A lightweight residual regional adapter and a linear prediction head are then optimized in this structured space. Across four downstream EEG tasks, MART achieves the best average Macro-F1 in our controlled adaptation comparison while updating only 0.108%–0.501% of the model parameters. Overall, MART reframes EEG representation adaptation as a problem of not only how to update a representation model, but also how its representations should be exposed to downstream adaptation.

1 INTRODUCTION

Electroencephalography (EEG) supports a broad range of applications, including affective computing, motor imagery, sleep staging, and clinical analysis. Recent EEG representation models increasingly aim to learn reusable representations that can be transferred across downstream tasks rather than requiring a separate task-specific representation model for each application (Yang et al., 2023; Jiang et al., 2024; El Ouahidi et al., 2025; Yang et al., 2026). This shift makes downstream adaptation increasingly important for transferring EEG representations across tasks. However, current adaptation strategies commonly operate directly on the representation interface exposed by the backbone, using linear prediction heads, lightweight adapters, low-rank updates, or other parameter-efficient mechanisms (Houlsby et al., 2019; Li & Liang, 2021; Zhang et al., 2023).

This raises a fundamental question for EEG adaptation. The representation interface exposed by the backbone may not be the most suitable space for downstream adaptation. Although EEG representation models often expose features as flat hidden states, these representations originate from structured channel-time measurements whose availability depends on the recording montage and dataset. Direct adaptation on flat representations can therefore conceal recoverable channel-time provenance and provide no explicit indication of whether regional measurements are fully observed, partially covered, or unavailable. EEG adaptation should consequently consider not only how many parameters are updated, but also how the representation interface is organized before adaptation.

We study EEG adaptation from this adaptation-interface perspective. Our central hypothesis is that downstream adaptation can benefit from first constructing a structured adaptation space rather than directly optimizing over the representation interface exposed by the backbone. This view separates representation transferability from adaptation-interface suitability. A representation can remain transferable across tasks while the form in which it is exposed to downstream adaptation is still suboptimal. We therefore treat the adaptation interface itself as a design choice rather than a fixed output of the representation model.

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.

Based on this view, we propose MART, a Measurement-Aware Regional Transformation framework for EEG representation adaptation. Given representations with recoverable channel-time provenance, MART restores their channel-time organization and reparameterizes them into dataset-specific scalp-region representations. Coverage and availability descriptors are further introduced to represent regional observation conditions. A lightweight residual regional adapter and a linear prediction head are then optimized in the resulting structured regional adaptation space. In this way, MART transforms the exposed representation interface before downstream specialization while keeping task-specific adaptation lightweight.

We evaluate MART across four downstream EEG tasks covering affective valence classification, emotion recognition, motor imagery classification, and sleep-stage classification. Our experiments first examine whether MART provides effective adaptation across compatible fixed-backbone EEG representation models and whether it remains competitive with generic parameter-efficient and flat-feature adaptation methods. We then conduct capacity-controlled analyses to test whether the observed gains can be explained by increased downstream modeling capacity alone. Finally, we evaluate robustness under channel degradation and analyze model sensitivity to structured regional removal to examine how the constructed regional adaptation space behaves under incomplete measurements. Together, these experiments assess not only downstream performance and parameter efficiency, but also whether transforming the adaptation interface provides benefits beyond simply increasing task-specific model capacity.

Our contributions are summarized as follows.

- We formulate EEG representation adaptation as an adaptation-interface construction problem. Rather than assuming that the flat representation interface exposed by a fixed backbone is already suitable for downstream adaptation, we study whether this interface should first be transformed to preserve recoverable channel-time provenance and regional measurement conditions.
- We propose MART, a lightweight measurement-aware regional transformation framework that constructs a structured regional adaptation space before downstream adaptation. MART recovers channel-time organization, reparameterizes flat representations into anatomical regional representations, and explicitly represents regional coverage and availability before lightweight regional adaptation.
- We provide controlled empirical evidence for structured adaptation-space construction across heterogeneous EEG downstream tasks. MART achieves strong performance with only 0.108%–0.501% trainable parameters across compatible representation models, while analyses under capacity variation, channel degradation, and structured regional removal further characterize when and how the constructed adaptation space benefits downstream adaptation.

2 Related Work

Task-specific EEG decoding and generalization. Early EEG decoders such as EEGNet (Lawhern et al., 2018), EEGConformer (Song et al., 2023), and ATCNet (Altaheri et al., 2023) learn task-specific representations directly for supervised EEG decoding, while cross-subject methods further address generalization under inter-subject variability (Wang et al., 2024b). These approaches optimize task-specific models for a target decoding problem. MART studies a different setting in which the EEG representation model is fixed and the main question is how its exposed representation interface should be organized before lightweight downstream adaptation.

EEG representation and foundation models. Recent EEG representation and foundation models aim to learn transferable representations from large-scale or heterogeneous recordings, as illustrated by BIOT (Yang et al., 2023), LaBraM (Jiang et al., 2024), EEGPT (Wang et al., 2024a), CBraMod (Wang et al., 2025), LUNA (Döner et al., 2025), and REVE (El Ouahidi et al., 2025). These models make it possible to reuse learned EEG representations across downstream tasks rather than training a separate representation model for each task. Their downstream interfaces, however, differ substantially. Some retain identifiable channel-time organization in intermediate representations, whereas others expose pooled features, channel-unified representations, or model-specific latent structures. This variation makes the representation interface itself an important considera-

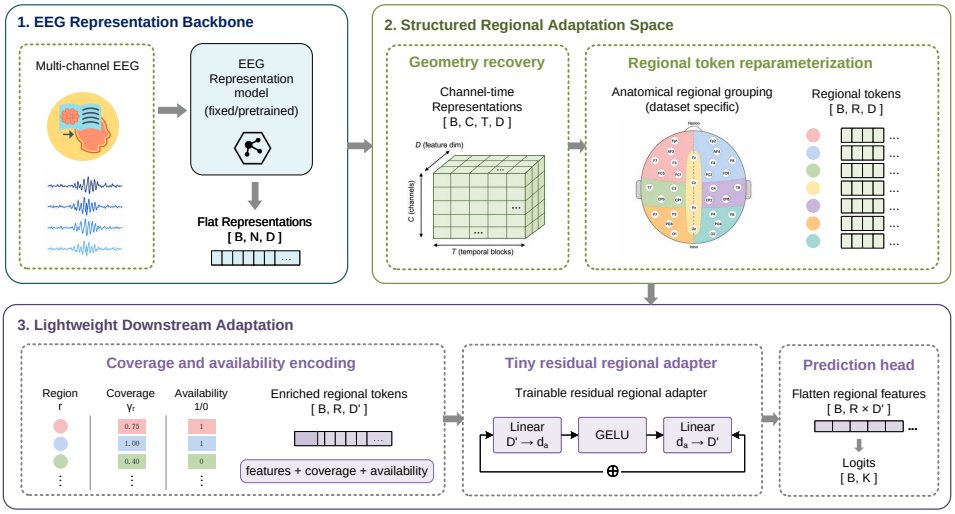

Figure 1: Overview of MART. MART transforms the exposed EEG representation interface into a structured regional adaptation space before lightweight downstream adaptation. For representations with recoverable channel-time provenance, it restores channel-time organization, constructs anatomical regional representations, incorporates regional coverage and availability information, and performs lightweight adaptation in the resulting structured space.

tion for downstream adaptation. MART builds on this distinction by considering representation interfaces with verifiable channel-time provenance and transforming them into a structured regional adaptation space.

Parameter-efficient and structure-aware adaptation. Parameter-efficient adaptation specializes an existing model by updating only a small subset of parameters or by introducing lightweight trainable components. Representative approaches include low-rank weight updates, residual adapters, and prompt- or prefix-based conditioning (Houlsby et al., 2019; Li & Liang, 2021; Hu et al., 2022; Liu et al., 2024). These methods differ in how adaptation is parameterized, but they generally operate on the representation interface provided by the original model. MART addresses a complementary question. Instead of asking only which lightweight parameterization should be optimized, it asks whether the exposed representation should first be transformed before downstream adaptation. MART therefore combines interface transformation with lightweight adaptation in the resulting structured regional space.

3 Method

3.1 Overview and Adaptation-Space Formulation

MART applies to EEG representation interfaces with recoverable channel-time provenance and transforms them into a structured regional adaptation space before lightweight downstream adaptation. Figure 1 illustrates the overall framework. A fixed EEG representation model first maps a multichannel EEG segment to an exposed representation. MART reorganizes this representation according to its verified channel-time structure and reparameterizes it into anatomical regional representations. Coverage and availability descriptors are then introduced to represent the observation condition of each region. Lightweight downstream adaptation is performed in the resulting structured space using a residual regional adapter and a prediction head. We do not infer geometry from token count, replicate pooled features across channels, or introduce pseudo-channel mappings for unsupported interfaces.

This pipeline can be viewed formally as a transformation of the downstream adaptation interface. Standard linear probing and generic parameter-efficient adaptation typically operate directly on the representation interface exposed by the backbone

$$o = h_\phi(Z), \qquad Z = F_\theta(X), \tag{1}$$

where X denotes an EEG input segment, F_θ denotes the EEG representation model with parameters θ, Z denotes the exposed representation, and h_ϕ denotes the downstream predictor with trainable parameters ϕ. The output o denotes the task prediction.

MART instead introduces an interface transformation before downstream adaptation

$$o = h_\phi\big(T_d(Z, M_d)\big), \tag{2}$$

where d indexes the downstream dataset, M_d denotes its montage and regional measurement information, and T_d denotes the dataset-specific interface transformation that maps the exposed representation into a structured regional adaptation space. The role of T_d is not to reconstruct missing EEG signals or infer unobserved neural activity. It reorganizes information that is already recoverable from the representation interface and makes regional measurement conditions explicit to the downstream adapter.

The exposed representation has the form

$$Z \in \mathbb{R}^{B\times N\times D}, \tag{3}$$

where B denotes the batch size, N denotes the number of exposed representation tokens, and D denotes the representation dimension. MART is applicable when N admits a verified mapping to the channel-time organization defined by the representation model.

3.2 Channel-Time Geometry Recovery

Compatible EEG representation models may expose intermediate features as flat token sequences even when these features retain a verified correspondence to the underlying channel-time measurement structure. MART uses this correspondence to reorganize the exposed representation before regional reparameterization. For compatible interfaces, the number of representation tokens satisfies $N = CT$, where C denotes the number of EEG channels and T denotes the number of temporal representation blocks or patches. Given a verified channel-major token ordering, the flat representation is reshaped as

$$\widetilde{Z}_{b,c,t,:} = Z_{b,cT+t,:}, \qquad \widetilde{Z} \in \mathbb{R}^{B\times C\times T\times D}. \tag{4}$$

Here, T denotes the temporal resolution of the representation rather than the sampling length of the raw EEG signal. Non-channel tokens, such as CLS tokens, are removed before this reorganization when present.

This operation does not infer spatial or temporal structure from token count alone. It only restores a channel-time organization that is already specified by the representation model and verified from its interface. The resulting channel-time representation $\widetilde{Z}$ provides the structured input for regional reparameterization.

3.3 Regional Representation Reparameterization

After recovering the channel-time organization, MART reparameterizes the exposed representation into a structured regional adaptation space. This transformation changes the unit of downstream adaptation from individual flat representation tokens to anatomical regional representations. The region definition is dataset-specific and follows the available electrode montage rather than assuming a universal partition across datasets.

Let $\mathcal{R}^{(d)} = \{r_1, r_2, \ldots, r_{R_d}\}$ denote the anatomical region set for dataset d, where R_d is the number of regions. Given $\widetilde{Z} \in \mathbb{R}^{B\times C\times T\times D}$, MART first aggregates the observed channels assigned to each region. For region r, let $\mathcal{C}_r^{\mathrm{obs}}$ denote the set of currently observed channels associated with that region. When $|\mathcal{C}_r^{\mathrm{obs}}| > 0$, regional channel aggregation is defined as

$$u_{b,r,t,:} = \frac{1}{|\mathcal{C}_r^{\mathrm{obs}}|} \sum_{c\in\mathcal{C}_r^{\mathrm{obs}}} \widetilde{Z}_{b,c,t,:}, \qquad u_{b,r,t,:} \in \mathbb{R}^D. \tag{5}$$

The temporal representations within each region are then aggregated as

$$h_{b,r} = \frac{1}{T}\sum_{t=1}^{T} u_{b,r,t,:}, \qquad h_{b,r} \in \mathbb{R}^D. \tag{6}$$

Stacking the regional representations yields

$$H = \{h_{b,r}\}_{r=1}^{R_d}, H \in \mathbb{R}^{B\times R_d\times D}. \tag{7}$$

This reparameterization forms the structural basis of the regional adaptation space used by the downstream components. It reduces the exposed channel-time representation to a compact set of region-level representations while retaining the correspondence between each representation and its anatomical measurement region. The purpose is to provide a structured adaptation interface rather than to estimate the underlying neural source activity.

During missing-channel or structured region-removal evaluation, a region may contain no observed channels. In this case, its regional representation is set to zero. The measurement descriptors introduced in the next subsection explicitly indicate whether that region is observed and how completely it is covered. No missing EEG signal or unobserved neural activity is reconstructed.

3.4 Coverage and Availability Encoding

MART augments each regional representation with deterministic descriptors of its measurement condition. For region r, the coverage ratio is defined as

$$\gamma_r = \frac{|\mathcal{C}_r^{\text{obs}}|}{|\mathcal{C}_r^{\text{ref}}|}, \tag{8}$$

where $\mathcal{C}_r^{\text{ref}}$ denotes the reference channel set of region r, and $\mathcal{C}_r^{\text{obs}}$ denotes its currently observed subset. The availability indicator is defined as

$$a_r = \mathbb{I}\left(|\mathcal{C}_r^{\text{obs}}| > 0\right). \tag{9}$$

Thus, γ_r measures how much of the reference region is observed, while a_r indicates whether the region has any observed channel.

The descriptors are concatenated with the regional representation

$$\bar{h}_{b,r} = [h_{b,r}; \gamma_r; a_r], \qquad \bar{h}_{b,r} \in \mathbb{R}^{D'}, \qquad D' = D + 2. \tag{10}$$

Stacking all regions gives $\bar{H} \in \mathbb{R}^{B\times R_d\times D'}$. This encoding makes regional coverage and availability explicit in the adaptation interface without reconstructing missing measurements.

3.5 Lightweight Residual Regional Adapter

After regional reparameterization and measurement encoding, MART performs lightweight adaptation in the structured regional space while keeping the EEG representation model fixed. A bottleneck MLP is applied independently to each enriched regional representation

$$A_\phi(x) = W_2\sigma(W_1x), \tag{11}$$

where $W_1 \in \mathbb{R}^{d_a\times D'}$, $W_2 \in \mathbb{R}^{D'\times d_a}$, $\sigma(\cdot)$ denotes GELU, and $d_a = \min(64, D')$. The regional representation is updated through a residual connection

$$h'_{b,r} = \bar{h}_{b,r} + A_\phi(\bar{h}_{b,r}), \qquad h'_{b,r} \in \mathbb{R}^{D'}. \tag{12}$$

Applying the same operation across regions gives $H' \in \mathbb{R}^{B\times R_d\times D'}$.

The adapter provides a low-capacity task-specific update within the constructed regional adaptation space rather than defining the adaptation space itself.

3.6 Prediction Head and Training Objective

The adapted regional representations are flattened according to the fixed dataset-specific region order

$$g = \text{Flatten}(H'), \qquad g \in \mathbb{R}^{B\times R_dD'}. \tag{13}$$

A linear classifier produces the downstream logits

$$o = W_cg + b_c, \qquad o \in \mathbb{R}^{B\times K}, \tag{14}$$

where K denotes the number of task classes. Training uses the standard cross-entropy objective

$$\mathcal{L} = \mathrm{CE}(o, y), \tag{15}$$

where y denotes the ground-truth label.

Overall, MART separates adaptation-space construction from downstream prediction, allowing task-specific optimization to operate on structured regional representations rather than directly on the exposed flat interface.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate MART on four public EEG benchmarks covering different tasks and recording configurations. DEAP (Koelstra et al., 2012) is used for binary valence classification, SEED (Zheng & Lu, 2015) for three-class emotion recognition, BCI Competition IV 2a (Tangermann et al., 2012) for four-class motor imagery classification, and Sleep-EDF (Kemp et al., 2000) for five-class sleep staging. Together, these datasets cover affective computing, motor imagery decoding, and sleep analysis.

Backbones and Baselines. We evaluate three groups of methods covering task-specific EEG decoding, frozen representation probing, and fixed-backbone adaptation. Task-specific baselines include EEGNet (Lawhern et al., 2018), EEGConformer (Song et al., 2023), and ATCNet (Altaheri et al., 2023). Frozen linear probing is evaluated with EEGPT (Wang et al., 2024a), LUNA (Döner et al., 2025), and CBraMod (Wang et al., 2025). MART is evaluated only with representation backbones whose exposed features retain recoverable channel-time provenance, including LaBraM (Jiang et al., 2024) and REVE (El Ouahidi et al., 2025). We do not apply MART to pooled or channel-unified interfaces for which this provenance cannot be verified. The main fixed-backbone comparison includes Linear Probe, Adapter (Houlsby et al., 2019), and LoRA+ (Hayou et al., 2024). A separate controlled comparison evaluates a broader set of generic parameter-efficient adaptation methods under the same backbone and downstream setting. This design controls for backbone differences and focuses the comparison on downstream adaptation strategies.

Training and Evaluation Protocol. For representation adaptation experiments, the EEG representation backbone remains fixed and only the adaptation module and prediction head are optimized. Task-specific EEG decoders are trained end to end using the same downstream data splits. All methods are evaluated with Accuracy and Macro-F1. Macro-F1 is used as the primary metric because the evaluated tasks include multi-class prediction and imbalanced class distributions. Unless otherwise specified, results are reported as mean and standard deviation over three random seeds.

4.2 Overall Comparison

Table 1 compares task-specific EEG decoders, frozen representation probing, and compatible fixed-backbone adaptation across four downstream tasks. Linear probing provides a basic downstream readout of fixed EEG representations, while MART evaluates whether transforming a compatible representation interface before lightweight adaptation provides additional benefits.

Task-specific decoders remain strong on several datasets but require end-to-end optimization for each downstream task. Frozen linear probing shows substantial variation across representation models and datasets, indicating that the representation interface exposed by an EEG model can substantially affect downstream prediction.

On compatible backbones, MART improves over Linear Probe in Macro-F1 across all available settings. It achieves the highest Macro-F1 within the compatible adaptation block on three of four LaBraM tasks and two of three available REVE tasks. Adapter performs better on LaBraM with SEED and REVE with Sleep-EDF, indicating that MART does not dominate every backbone and task combination. Nevertheless, MART remains competitive while updating only 0.501% of LaBraM parameters and 0.108% of REVE parameters.

Overall, these results establish MART as an effective lightweight adaptation approach for EEG representation interfaces with recoverable channel-time provenance. The controlled analyses in the

Table 1: Overall EEG adaptation results across four downstream tasks. Each dataset reports Accuracy and Macro-F1 in percent. Best and second-best results within each backbone-specific adaptation block are indicated by **bold** and underlined values, respectively. Unavailable or incompatible settings are marked as –.

Setting	Backbone	Method	DEAP Acc. / F1 (%)	SEED Acc. / F1 (%)	BCI IV 2a Acc. / F1 (%)	Sleep-EDF Acc. / F1 (%)	Train. Ratio
Task-specific	EEGNet (JNE 2018)		52.7±8.9 / 49.0±9.7	44.5±5.9 / 43.7±6.1	50.0±6.7 / 48.4±7.1	56.9±23.4 / 30.8±14.9	100.00%
	EEGConformer (TNSRE 2023)		54.5±10.4 / 35.2±5.0	45.8±6.7 / 41.6±7.5	44.7±6.5 / 43.5±7.0	58.7±24.0 / 28.7±12.0	100.00%
	ATCNet (TII 2023)		56.3±9.2 / 42.8±7.7	46.0±7.7 / 45.3±7.9	47.1±5.8 / 45.6±6.1	63.0±22.2 / 33.9±11.3	100.00%
Foundation + Linear Probe	EEGPT (NeurIPS 2024)		47.8±4.2 / 41.4±7.4	–	28.0±3.3 / 24.7±7.0	83.9±0.5 / 51.7±1.4	0.007%
	LUNA (NeurIPS 2025)		48.3±1.4 / 35.0±4.8	31.1±3.8 / 18.2±2.7	25.2±0.5 / 12.0±3.0	65.2±0.2 / 15.8±0.0	0.013%
	CBraMod (ICLR 2025)		46.7±0.0 / 31.8±0.0	35.6±3.1 / 23.0±5.7	30.4±0.3 / 21.7±2.4	82.9±0.2 / 49.1±0.1	0.014%
Compatible adaptation	LaBraM (ICLR 2024)	Linear Probe	47.5±2.5 / 35.2±5.4	36.7±3.3 / 31.7±0.8	26.4±1.6 / 24.9±1.8	77.4±1.8 / 54.2±3.8	0.012%
		Adapter	47.5±0.0 / 32.2±0.0	**46.7±3.8 / 39.7±7.5**	26.5±1.2 / 21.3±2.1	**83.1±2.4** / 49.7±2.7	5.149%
		LoRA+	50.8±2.9 / 33.7±1.3	31.1±3.8 / 17.3±1.1	25.2±0.4 / 10.7±1.3	69.8±6.3 / 22.0±8.1	5.026%
		MART (ours)	**51.1±6.9 / 48.1±9.0**	40.0±6.7 / 36.9±8.5	**34.8±2.2 / 33.7±2.1**	81.1±0.2 / **55.8±2.7**	0.501%
	REVE (NeurIPS 2025)	Linear Probe	**51.7±1.4** / 35.4±1.7	–	33.3±0.3 / 27.8±2.7	84.9±0.1 / 54.8±1.8	0.003%
		Adapter	47.5±0.0 / 32.2±0.0	–	30.4±0.5 / 23.6±2.6	**86.3±1.1 / 58.4±3.5**	7.866%
		LoRA+	49.2±2.9 / 32.9±1.3	–	24.9±0.1 / 12.2±3.8	65.3±0.0 / 15.8±0.0	0.778%
		MART (ours)	50.0±2.5 / **44.2±7.3**	–	**45.4±4.1 / 41.0±5.2**	80.9±4.2 / 57.1±1.7	0.108%

Table 2: Controlled comparison with parameter-efficient adaptation methods using the fixed LaBraM backbone. Each dataset reports Accuracy and Macro-F1 in percent. Best and second-best results are indicated by **bold** and underlined values, respectively.

Method	DEAP Acc. / F1 (%)	SEED Acc. / F1 (%)	BCI IV 2a Acc. / F1 (%)	Sleep-EDF Acc. / F1 (%)	Trainable Params
Linear Probe	47.5±2.5 / 35.2±5.4	28.9±9.7 / 24.6±9.2	26.4±1.6 / 24.9±1.8	75.6±2.6 / 31.5±3.0	704
Adapter (ICML 2019)	47.5±0.0 / 32.2±0.0	**46.7±3.8 / 39.7±7.5**	26.5±1.2 / 21.3±2.1	**83.1±2.4** / 49.7±2.7	315,872
Prefix Tuning (ACL 2021)	47.5±0.0 / 32.2±0.0	35.6±3.8 / 30.3±4.0	23.1±0.4 / 22.4±0.8	80.3±0.4 / 33.0±0.3	2,304
AdaLoRA (ICLR 2023)	45.8±2.9 / 32.5±0.5	29.6±3.4 / 20.7±3.1	24.4±2.0 / 21.9±2.4	79.1±1.7 / 33.8±2.4	116,048
VeRA (ICLR 2024)	**51.7±3.8** / 40.2±7.3	40.7±11.0 / 35.8±14.1	24.1±1.1 / 22.2±0.4	78.5±1.2 / 33.2±1.2	8,000
FourierFT (ICML 2024)	45.8±2.9 / 34.6±4.2	32.6±1.3 / 29.2±2.3	25.3±2.1 / 24.2±1.8	77.5±2.7 / 32.1±2.1	1,472
DoRA (ICML 2024)	50.0±4.3 / 40.4±11.2	42.2±9.7 / 37.1±9.1	26.0±0.6 / 24.8±1.7	78.7±0.8 / 36.3±4.9	84,704
LoRA+ (ICML 2024)	50.8±2.9 / 33.7±1.3	31.1±3.8 / 17.3±1.1	25.2±0.4 / 10.7±1.3	69.8±6.3 / 22.0±8.1	307,904
MART (ours)	51.1±6.9 / **48.1±9.0**	40.0±6.7 / 36.9±8.5	**34.8±2.2 / 33.7±2.1**	81.1±0.2 / **55.8±2.7**	29,307

following sections further examine whether the observed gains are primarily associated with structured regional adaptation-space construction rather than increased adaptation capacity.

4.3 Comparison with Generic Parameter-Efficient Adaptation

Table 2 compares MART with generic parameter-efficient adaptation methods using the same fixed LaBraM backbone and downstream protocol. This controlled setting keeps the representation model and evaluation protocol unchanged across methods, allowing the comparison to focus on different adaptation strategies. The baselines include Linear Probe, Adapter (Houlsby et al., 2019), Prefix Tuning (Li & Liang, 2021), AdaLoRA (Zhang et al., 2023), VeRA (Kopiczko et al., 2024), FourierFT (Gao et al., 2024), DoRA (Liu et al., 2024), and LoRA+ (Hayou et al., 2024). Additional recent parameter-efficient methods are reported in Appendix.

MART achieves the highest Macro-F1 on DEAP, BCI IV 2a, and Sleep-EDF, while remaining close to the strongest methods on SEED. It uses 29,307 trainable parameters, fewer than Adapter, AdaLoRA, DoRA, and LoRA+, although several generic methods use fewer trainable parameters. The results show that MART remains competitive with a diverse set of parameter-efficient adaptation strategies under the same representation backbone. Linear probing remains a strong baseline in several settings, but its effectiveness is task and backbone dependent. MART shows a more consistent performance profile across affective decoding, motor imagery, and sleep staging, supporting the broader applicability of structured regional adaptation across heterogeneous EEG tasks.

Table 3: Capacity controlled ablation of adaptation space construction. We compare flat and regional variants using linear prediction, nonlinear MLP prediction, and lightweight residual adaptation. The regional MLP closely matches the trainable parameter count of Full MART, while the flat MLP serves as a nonlinear capacity control. Results are reported as mean ± standard deviation over three seeds.

Variant	DEAP			BCI IV 2a		
	Params	Acc.	Macro-F1	Params	Acc.	Macro-F1
Flat Linear	64,002	53.3±6.7	50.3±7.2	88,004	30.3±2.9	29.3±3.4
Flat MLP	32,005	53.3±0.0	34.8±0.0	22,009	25.0±0.0	10.0±0.0
Flat Adapter	26,266	46.7±3.3	37.3±5.0	26,668	28.3±1.5	23.3±6.0
Regional Token + Linear	2,802	**55.6±6.9**	**55.0±7.1**	4,804	**32.1±1.2**	**31.9±1.3**
Regional Token + MLP	29,465	46.7±8.8	43.3±6.8	31,334	29.2±4.0	19.3±9.3
Full MART	28,952	53.3±14.5	50.1±16.1	30,974	30.9±0.6	29.5±2.0

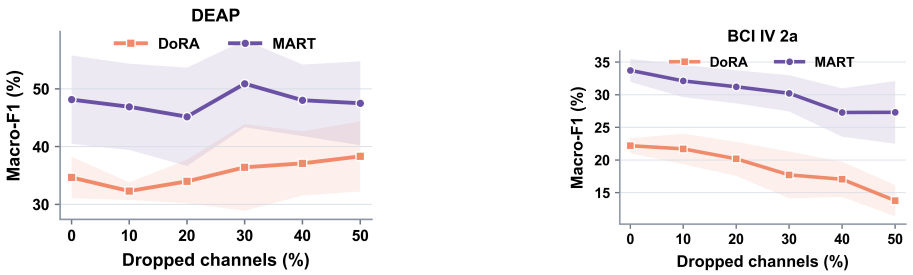

Figure 2: Robustness under random test-time channel dropping. Macro-F1 is reported across channel drop ratios from 0% to 50% and averaged over training seeds and repeated random masks.

Generic parameter-efficient methods primarily change how a limited set of parameters is updated while retaining the original adaptation interface. MART instead transforms the exposed representation into a structured regional adaptation space before lightweight optimization.

4.4 Capacity-Controlled Analysis of the Adaptation Space

We examine whether the gains of MART can be explained by increased downstream adaptation capacity rather than by structured regional reparameterization. To this end, we compare flat and regional representations using linear prediction, nonlinear MLP prediction, and lightweight residual adaptation. The regional MLP is configured to closely match the trainable parameter count of Full MART, while the flat MLP serves as a nonlinear capacity control because exact parameter matching is not always feasible under the high-dimensional flattened input.

Table 3 shows a clear difference between increasing downstream capacity and changing the adaptation space. For flat representations, neither the nonlinear MLP nor the residual adapter reproduces the performance obtained after regional reparameterization. On DEAP and BCI IV 2a, Regional Token + Linear achieves the highest Macro-F1 with only 2.8K and 4.8K trainable parameters, respectively, and also outperforms Full MART. Increasing model capacity within the regional space does not provide consistent additional gains, as Regional Token + MLP remains below the corresponding linear variant despite closely matching the parameter count of Full MART.

These results show that the observed improvements cannot be explained by increased downstream capacity alone and provide evidence that changing the adaptation space itself affects downstream performance.

4.5 Robustness and Regional Analysis

Random channel drop robustness. We evaluate MART under test-time channel dropping as a stress test of adaptation under incomplete observations. Models are trained with the complete mon-

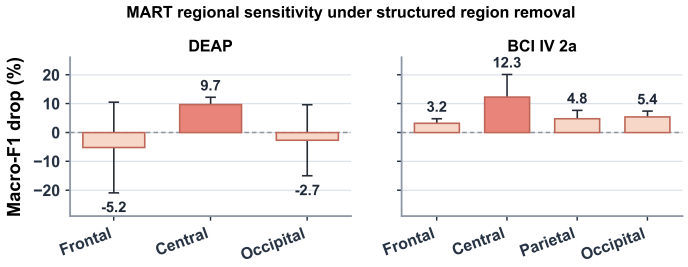

Figure 3: Model-level regional sensitivity under structured region removal. Bars report the Macro-F1 change relative to the complete montage after removing individual anatomical regions.

tage and evaluated without retraining after randomly removing 0% to 50% of EEG channels. For each drop ratio, results are averaged over training seeds and multiple random channel-drop masks. MART recomputes the regional representations and measurement descriptors from the remaining observed channels.

Figure 2 reports results on DEAP and BCI IV 2a. Compared with DoRA, MART maintains a more favorable degradation profile across most channel-drop levels on both datasets. The curves are not strictly monotonic because different random masks remove different subsets of measurements, but the overall trend indicates that structured regional adaptation remains effective as channel observations become incomplete. An internal missingness ablation in Appendix further shows that regional reparameterization remains the dominant factor under channel dropping, while the additional benefit of measurement-aware residual adaptation varies across tasks.

Regional sensitivity. We further analyze MART under structured regional removal. Models are trained on the complete montage and evaluated under test-time region removal without retraining. For each selected predefined scalp region, all observed channels assigned to that region are removed from the test input. MART then recomputes regional representations together with coverage and availability indicators from the remaining channels. Unlike random channel dropping, this setting removes coherent regional measurements rather than isolated channels.

Figure 3 reports the resulting Macro-F1 change relative to the complete montage. On BCI IV 2a, removal of the central region produces the largest degradation, whereas DEAP shows a more distributed sensitivity pattern across regions. These results characterize task-dependent model sensitivity to coherent regional measurement loss within the structured adaptation space and should be interpreted as model-level dependence rather than causal neural-region importance.

5 Conclusion and Future Work

We present MART, a measurement-aware regional transformation framework for EEG representation adaptation. Rather than directly adapting the representation interface exposed by an EEG backbone, MART first constructs a structured regional adaptation space by recovering verified channel-time organization, forming scalp-region representations, and incorporating regional coverage and availability information before lightweight downstream adaptation. Across heterogeneous EEG tasks and compatible representation models, MART achieves strong downstream performance with a small fraction of trainable parameters. Controlled analyses further show that the observed gains cannot be explained by increased downstream capacity alone, providing evidence that the form of the adaptation space itself matters for downstream EEG transfer. Robustness and regional analyses also show that the structured space provides a natural interface for handling incomplete measurements and examining model-level dependence on regional observations.

These findings highlight the representation interface as an important design dimension for EEG representation transfer. Future work will extend MART to a broader range of representation interfaces, investigate more flexible regional constructions for varying montages and recording conditions, and explore structured adaptation for other physiological and sensor representations beyond EEG.

REFERENCES

Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman. Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. *IEEE Transactions on Industrial Informatics (TII)*, 19(2):2249–2258, 2023. doi: 10.1109/TII.2022.3197419.

Berkay Döner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. LUNA: Efficient and topology-agnostic foundation model for eeg signal analysis. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, pp. 70682–70708. Curran Associates, Inc., 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/66969a9e6bd7a26dfeccea7227178ca7-Paper-Conference.pdf.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A foundation model for EEG: Adapting to any setup with large-scale pretraining on 25,000 subjects. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. URL https://brain-bzh.github.io/reve/.

Ziqi Gao, Qichao Wang, Aochuan Chen, Zijing Liu, Bingzhe Wu, Liang Chen, and Jia Li. Parameter-efficient fine-tuning with discrete Fourier transform. In *Proceedings of International Conference on Machine Learning (ICML)*, volume 235, pp. 14884–14901, 2024.

Soufiane Hayou, Nikhil Ghosh, and Bin Yu. LoRA+: Efficient low rank adaptation of large models. In *International Conference on Machine Learning (ICML)*, 2024.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for NLP. In *International Conference on Machine Learning (ICML)*, pp. 2790–2799, 2019.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *International Conference on Learning Representations (ICLR)*, 2024.

Bob Kemp, Aeilko H. Zwinderman, Bert Tuk, Hilbert A. C. Kamphuisen, and Josef J. L. Oberye. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering (IEEE TBME)*, 47(9):1185–1194, 2000. doi: 10.1109/10.867928.

Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing (IEEE TAC)*, 3(1):18–31, 2012. doi: 10.1109/T-AFFC.2011.15.

Dawid J. Kopiczko, Tijmen Blankevoort, and Yuki M. Asano. VeRA: Vector-based random matrix adaptation. In *International Conference on Learning Representations (ICLR)*, 2024.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A compact convolutional network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018. doi: 10.1088/1741-2552/aace8c.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Annual Meeting of the Association for Computational Linguistics*, pp. 4582–4597, 2021.

Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. DoRA: Weight-decomposed low-rank adaptation. In *International Conference on Machine Learning (ICML)*, 2024.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. EEG conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering (IEEE TNSRE)*, 31:710–719, 2023. doi: 10.1109/TNSRE.2022.3230250.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J. Miller, Gernot R. Mueller-Putz, Guido Nolte, Gert Pfurtscheller, Hubert Preissl, Gerwin Schalk, Alois Schlögl, Carmen Vidaurre, Stephan Waldert, and Benjamin Blankertz. Review of the BCI competition IV. *Frontiers in Neuroscience*, 6:55, 2012. doi: 10.3389/fnins.2012.00055.

Guagnyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. EEGPT: Pretrained transformer for universal and reliable representation of eeg signals. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 39249–39280. Curran Associates, Inc., 2024a. doi: 10.52202/079017-1239. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/4540d267eeec4e5dbd9dae9448f0b739-Paper-Conference.pdf.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *International Conference on Learning Representations (ICLR)*, 2025.

Yiming Wang, Bin Zhang, and Yujiao Tang. DMMR: cross-subject domain generalization for eeg-based emotion recognition via denoising mixed mutual reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024b. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i1.27819. URL https://doi.org/10.1609/aaai.v38i1.27819.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Cross-data biosignal learning in the wild. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Liuyin Yang, Qiang Sun, Ang Li, and Marc M. Van Hulle. Are EEG foundation models worth it? comparative evaluation with traditional decoders in diverse BCI tasks. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=5Xwm8e6vbh.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Nikos Karampatziakis, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *International Conference on Learning Representations (ICLR)*, 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. doi: 10.1109/TAMD.2015.2431497.