UNIPOPT: WHERE DOES INTRACRANIAL EEG HELP A BRAIN-SIGNAL FOUNDATION MODEL?

Anonymous authors
Paper under double-blind review

ABSTRACT

Intracranial EEG (iEEG) is often described as an untapped resource for brain-signal foundation models, yet whether it improves decoding has not been established: the one prior report of a gain from joint EEG–iEEG pre-training used a model that carries no electrode coordinates. We present UNIPOPT, a Population Transformer that ingests scalp EEG and iEEG through one interface: a frozen single-channel encoder turns every electrode into a token, a spherical-coordinate and modality positional encoding places it, and the aggregator is pre-trained on both modalities with next-segment and replaced-token objectives. On the official evaluation splits of two widely used clinical EEG corpora UNIPOPT improves on existing methods trained on the same recordings; on the external Brain4FMs benchmark it obtains the best mean rank on the scalp-EEG leaderboards and the top accuracy on the intracranial ones. Decomposing the gain on the same external benchmark, we find that the full model beats every ablated variant, that the gain enters through iEEG in the aggregator's pre-training corpus, and that the spherical encoding is what lets the model exploit electrode montages it never saw during pre-training. Beyond benchmark accuracy, we analyse how decoding scales with the amount of fine-tuning data per subject and show that most of the gain from additional electrodes arrives through the spherical coordinates.

1 INTRODUCTION

Scalp EEG and intracranial EEG (iEEG) record the same electrical activity of the brain from different distances. Scalp EEG is plentiful and cheap but attenuated by skull and scalp; iEEG records directly from cortex and depth structures with high signal-to-noise ratio, but it is scarce, expensive and confined to clinical monitoring. Foundation models for brain signals have so far grown by adding scalp hours (Yang et al., 2023; Jiang et al., 2024; Wang et al., 2025; El Ouahidi et al., 2025). Two facts are already established: pre-training a channel-independent model jointly on both modalities can transfer between them (Yuan et al., 2024), and giving a scalp-EEG model electrode coordinates helps it generalise across montages (El Ouahidi et al., 2025). Our goal is the step that neither takes: to fuse the two modalities in one model through a coordinate interface that both can share, spherical coordinates about the head centre, and to show that the intracranial recordings then improve decoding rather than merely coexist with the scalp data.

The obstacle is aggregation, not embedding. Single-channel temporal encoders have become strong (Wang et al., 2023; Yue et al., 2022; Jiang et al., 2024), which suggests the recipe that the Population Transformer (PopT) (Chau et al., 2025) established for iEEG: embed every channel on its own, then learn to aggregate the embeddings across channels. For one modality the hard part is already the variable electrode layout; for two, the layouts are not even of the same kind. Scalp montages are standard and can be placed on a template, whereas every implant has its own contacts at its own positions, on the cortical surface or along depth shafts, and the two modalities differ in reference scheme, amplitude and bandwidth as well. In practice this forces a choice that costs accuracy. Scalp datasets are cut down to the electrodes a model was pre-trained on, which throws away most of a dense montage. The opposite shortcut, feeding every channel to a model built for a fixed set, can hurt just as much when the fine-tuning data are too few for the model to learn the positions it never saw (Sec. 5). Intracranial data suffer more, because with no shared coordinate frame each patient needs a model calibrated on their own recording, and the hours of iEEG that already exist stay outside every scalp-EEG model. The question of how two modalities enter one model is

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.

therefore a practical one, and it decides whether iEEG can be a resource for scalp-EEG decoding at all.

Existing approaches stop short of this. Channel-independent joint pre-training treats channels as an unordered bag, so the model cannot know where a token came from, and its reported gain was measured against a single-modality corpus of half the size, which leaves modality and data volume entangled (Yuan et al., 2024). Coordinate-aware models cover one modality each: REVE places scalp electrodes at (x, y, z, t) (El Ouahidi et al., 2025), PopT places intracranial contacts at integer xyz (Chau et al., 2025). Cartesian coordinates are also the wrong language for two modalities: a scalp electrode and a depth contact at the same angular position differ only in radius, which xyz hides.

We present **UNIPOPT**, which keeps PopT's division of labour, a frozen single-channel TS2Vec encoder (Yue et al., 2022) and a learned aggregator with PopT's next-segment and replaced-token objectives, and changes three things: iEEG enters the bank the encoder is trained on, the positional encoding is built from spherical coordinates, depth and a modality code in place of Cartesian xyz, and the aggregator is pre-trained on a bimodal corpus in place of a scalp-only one (Fig. 1). The frozen encoder keeps training lightweight, and because iEEG can be added to the encoder's training bank and to the aggregator's corpus independently, each level is compared with its scalp-only counterpart at the same compute.

Empirically, on the official TUAB and TUEV splits UNIPOPT outperforms existing methods trained on the same recordings, and on the external Brain4FMs benchmark (Shen et al., 2026) it obtains the best mean rank on the scalp-EEG leaderboards and the top accuracy on the intracranial ones, with a pre-training corpus an order of magnitude smaller than the largest joint corpus. Running the ablations on both benchmarks shows where the gain comes from: iEEG helps through the aggregator's pre-training corpus, a little on the official splits and more on the external boards, and not through the encoder's training bank; the spherical encoding is what lets the model exploit electrode montages it never saw in pre-training, turning additional channels into a gain about four times what the Cartesian variant obtains (Sec. 5). Beyond accuracy, our recording-length analysis shows that a dense montage keeps improving with the data recorded per subject where the standard 19-electrode input tops out, so that extra channels pay off with enough data and the pre-training montage remains the safer choice on small datasets.

Our main contributions are:

- UNIPOPT, a single model that serves both 19–64-channel scalp EEG and 36–195-contact iEEG, built on a spherical-coordinate interface the two modalities share: template scalp electrodes and patient-specific intracranial contacts enter through the same positional encoding, so one checkpoint covers either modality and can exploit electrode montages it never saw in pre-training.
- A bimodal pre-training recipe in which iEEG improves scalp-EEG decoding: intracranial recordings enter the aggregator's pre-training corpus through the shared interface, and a decomposition on both benchmarks localises the gain to that corpus rather than to the encoder bank, so the hours of clinical iEEG that already exist become a pre-training resource for scalp EEG.
- Evaluation on both modalities: on the official TUAB and TUEV splits UNIPOPT outperforms existing methods trained on the same recordings, and on ten Brain4FMs boards run under the benchmark's own protocol it obtains the best mean rank on the scalp-EEG leaderboards and the top accuracy on the intracranial ones, with the one choice on our side, the montage, made on validation folds.
- Analyses of decoding versus fine-tuning data recorded per subject and of montage, showing that the 64-channel advantage grows with recording length and is about four times larger under the spherical encoding than under the Cartesian one.

2 RELATED WORK

Scalp and intracranial models. Self-supervised scalp-EEG models moved from fixed montages (Kostas et al., 2021; Chien et al., 2022) to topology-agnostic designs (Yi et al., 2023; Döner et al., 2025; El Ouahidi et al., 2025), and today's foundation models (Yang et al., 2023; Jiang et al., 2024; Wang et al., 2025; 2024; Jiang et al., 2025; Ma et al., 2026) are pre-trained on scalp recordings alone. Intracranial models (Wang et al., 2023; Zhang et al., 2023; Han et al., 2025) likewise stay

within their modality, and MBrain trains separate EEG and SEEG models (Cai et al., 2023). Treating each sensor as a token that carries its own identity was established for spiking populations (Azabou et al., 2023; Ye et al., 2025), and PopT (Chau et al., 2025) brought it to iEEG by aggregating a coordinate-identified electrode set over a frozen single-channel encoder such as TS2Vec (Yue et al., 2022). None of these models reads both modalities, so the intracranial recordings that exist cannot inform a scalp-EEG model; we extend PopT's electrode-set interface to both.

Joint EEG–iEEG pre-training. BrainWave (Yuan et al., 2024) pre-trains one channel-independent model on EEG and iEEG without electrode coordinates. We place scalp and intracranial electrodes in one coordinate frame and pre-train on both.

Spatial encodings. REVE encodes each channel by (x, y, z, t) (El Ouahidi et al., 2025) and PopT by an integer Cartesian lookup (Chau et al., 2025); spherical coordinates added to CBraMod give task-dependent effects without a consistent advantage (Yuce & Stober, 2026). Each of these encodings serves one modality, and Cartesian coordinates hide that a scalp electrode and a depth contact at the same angle differ only in radius. Our encoding is spherical and adds depth and modality, so that both kinds of electrode share one interface.

Evaluation, scaling and montage. Brain4FMs (Shen et al., 2026) evaluates EEG and iEEG foundation models under one subject-level cross-validation protocol. Fine-tuning curves by subject count exist only in AdaBrain-Bench (Wu et al., 2025), and transfer from few to many channels (Yi et al., 2023; Yang et al., 2023) and channel adaptation (Kokate et al., 2026) have been studied at a fixed data size. How the value of a denser montage depends on the fine-tuning data has not been measured; we evaluate under the Brain4FMs protocol and vary montage and recording length together.

3 Uni-Population Transformer Approach

Figure 1 overviews our Uni-Population Transformer (UniPopT) approach. The key ideas are: (1) to pre-train one model on scalp and intracranial recordings together; and (2) to place every electrode, on the scalp or inside the brain, in one spherical coordinate frame that both modalities share. We thereby aim at a single bimodal model in which intracranial recordings help scalp-EEG decoding and the spherical coordinates carry the model to montages it never saw in pre-training. Afterwards, one can fine-tune UniPopT on downstream decoding tasks, here the official TUAB and TUEV splits and the external Brain4FMs benchmark (Shen et al., 2026). In addition to offering strong decoding results, UniPopT can be ablated one change at a time, to locate the gain and to examine how iEEG benefits the model (Sec. 5).

Data. The pre-training corpus $\mathcal{D}$ joins the official training sets of TUAB and TUEV (Obeid & Picone, 2016; Harati et al., 2015) with five public iEEG datasets from NEMAR (Delorme et al., 2022), 1,136 h of scalp EEG and 691 h of iEEG in all; the official evaluation splits and every Brain4FMs subject are left out. We pre-train UniPopT and every ablation variant of Sec. 5 on this corpus, the EEG-only variants on its scalp part. Appendix A.1 lists the datasets.

Architecture. Consider a recording with channels $\mathcal{C} = \{1, \dots, N_c\}$, where channel i has modality $m_i \in \{\text{scalp}, \text{ECoG}, \text{SEEG}\}$ and MNI position $\mathbf{p}_i \in \mathbb{R}^3$, from a template for scalp electrodes and from the subject's electrode table for intracranial contacts. Let $x_i^t \in \mathbb{R}^T$ be its window starting at time t, after PopT's preprocessing, resampling to f_s and a per-window z-score ($T = f_s w$ for windows of w seconds); the model reads $X^t = \{x_i^t \mid i \in S\}$ for a channel subset $S \subseteq \mathcal{C}$ and a [CLS] token. Each window is embedded by a single-channel encoder $E : \mathbb{R}^T \to \mathbb{R}^{d_E}$, TS2Vec (Yue et al., 2022), one of PopT's EEG encoders: a dilated convolutional network trained with a hierarchical contrastive loss across time scales, whose max-pool over time embeds the window. We train E once on a bank $\mathcal{B}$ of $N_{\mathcal{B}}$ single-channel windows sampled at random from every recording of the pre-training corpus $\mathcal{D}$, iEEG more densely so that the bank keeps the corpus's scalp-to-iEEG ratio (EEG-only, at equal size, in the ablations), and then freeze it. The aggregator input is

$$H^t = \left\{ \mathrm{LN}\big(WE(x_i^t) + \mathrm{pos}(i)\big) \;\middle|\; i \in S \right\} \in \mathbb{R}^{|S| \times d}, \tag{1}$$

with a learned projection $W \in \mathbb{R}^{d \times d_E}$ and layer normalisation LN. PopT's $\mathrm{pos}(i)$ concatenates sinusoidal encodings of integer Cartesian coordinates and of the segment; we keep the mechanism

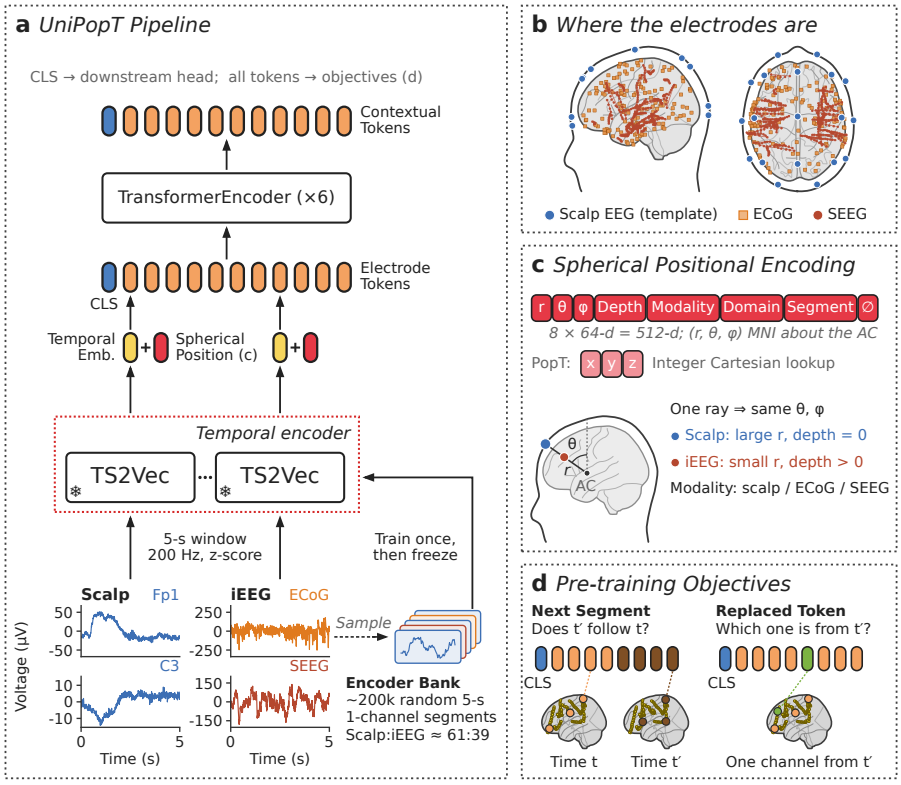


Figure 1: **Schematic of UniPopT.** (a) The pipeline, bottom to top: every electrode of either modality becomes one token of the single-channel encoder (Yue et al., 2022), which is trained once on the encoder bank and then frozen (snowflake; the red dotted region is what stays frozen). The temporal embedding (yellow) is added to the spherical positional embedding (red) to form the electrode tokens (orange), which, with a CLS token (blue), enter the trainable PopT aggregator (Chau et al., 2025) and come out as contextual tokens. Traces are real recordings, unfiltered, each on its own voltage scale. (b) Where the corpus electrodes are: SEEG depth shafts, ECoG contacts (squares) and the scalp template electrodes, on a side and a top view of the head. (c) The spherical positional encoding, the one architectural change to PopT: a scalp electrode and an intracranial contact on one ray share θ and φ and differ in r and depth. (d) PopT's two pre-training objectives: next segment (ensemble-wise, the orange and brown token sets from two segments) and replaced token (channel-wise, one green token from another segment).

and change the geometry:

$$\mathrm{pos}(i) = \big[\, e(r_i);\, e(\theta_i);\, e(\varphi_i);\, e(\delta_i);\, e(m_i);\, e(g_i);\, e(s_i);\, \mathbf{0} \,\big], \tag{2}$$

eight slots of $d/8$ dimensions, with $e(\cdot)$ PopT's sinusoidal encoding, $(r_i, \theta_i, \varphi_i)$ the spherical coordinates of $\mathbf{p}_i$ about the anterior commissure, δ_i the depth below the scalp (0 on the scalp), g_i the domain (scalp or intracranial) and s_i PopT's segment flag. A scalp electrode and an intracranial contact on one ray share (θ, φ) and differ only in r and δ, which the spherical form isolates (Fig. 1c). Where the geometry alone is close, as for ECoG contacts on the cortical surface beneath scalp electrodes, the modality code m_i still tells them apart.

Pre-training. We pre-train the aggregator A, PopT's transformer encoder, exactly as in PopT (Chau et al., 2025), with its ensemble-wise objective (the [CLS] token decides whether two channel ensembles are consecutive in time) and its channel-wise objective (each token decides whether its channel was swapped with activity from another time); the formulation is in Appendix A.2. No cross-modal term is added: iEEG reaches the scalp-EEG representation only through the shared E and A, and what changes from PopT is the data. The banks and the values of f_s, w, $N_{\mathcal{B}}$, d_E and d are in Appendix A.

Fine-tuning. Fine-tuning also follows PopT: a single linear layer maps the [CLS] output $\mathbf{y}_{\text{cls}} \in \mathbb{R}^d$ to class scores under a cross-entropy loss, A is trained end-to-end and E stays frozen; on Brain4FMs the benchmark's own head and loss replace the linear layer (Appendix A.3). Any channel that resolves in the template can enter S; whether a dataset is read at all such channels or at the pre-training montage is chosen on its validation folds (Sec. 4).

4 UniPopT Performance Evaluation

In-domain evaluation and results. We evaluate on the official TUAB and TUEV evaluation splits with 5 s windows (Table 1) and report, for UniPopT, the mean and standard deviation over 5 random fine-tuning seeds. At face value, UniPopT is ahead of BIOT on TUAB and in TUEV κ, although BIOT is pre-trained on a far larger corpus, and trails LaBraM-Base by 0.70 TUAB AUROC, with LaBraM further ahead on TUEV. On TUAB it also exceeds the published PopT result, obtained on the authors' own split, by 0.69, and P4, the PopT recipe we re-ran on our recordings, by 0.64. The comparisons with published numbers say little, however, because each rests on its own corpus: retrained on the same scalp recordings as P4, BIOT falls well below its published numbers, and below P4 on TUAB and in TUEV κ. The PopT recipe, by contrast, keeps its published TUAB level on the same recordings (P4 against the published PopT row), so the drop is not a property of our recordings. Most of UniPopT's lead over the retrained BIOT is therefore the PopT recipe's, and its standing against BIOT and LaBraM cannot be credited to our changes alone; what they add is the gap to P4, which we decompose in Sec. 5. The published numbers were also obtained under different splits and windows, so we next compare UniPopT with published models on a benchmark that runs every model under one protocol.

Table 1: TUAB abnormal-EEG detection and TUEV event classification on the official evaluation splits. Numbers are AUROC (TUAB), weighted F1 and Cohen's κ (TUEV), in percent, mean $\pm$ std over fine-tuning seeds. Tier (A) copies published numbers obtained on other corpora and other protocols; tier (B) is the same architecture retrained on exactly our scalp recordings; tier (C) is ours, with P4, the original PopT recipe run on the same scalp recordings, as the data-matched baseline.

Model	Pre-training	AUROC	Weighted F1	κ
(A)				
BIOT (Yang et al., 2023)	6 EEG corpora, ∼57,000 h	88.15	74.92	52.7
LaBraM-Base (Jiang et al., 2024)	∼2,500 h EEG	90.22	83.12	66.4
TS2Vec+PopT (Chau et al., 2025)	TUH, 21 ch	88.83	–	–
(B)				
BIOT (retrained)	TUH, ∼1,136 h	84.76	70.20	43.71
(C)				
UniPopT, 5 seeds	TUH + iEEG, ∼1,827 h	89.52 ± 0.10	71.83 ± 0.26	62.39 ± 0.50
P4 (PopT recipe), 3 seeds	TUH, ∼1,136 h	88.88 ± 0.25	70.49 ± 1.22	60.66 ± 1.91

External evaluation and results. Brain4FMs (Shen et al., 2026) runs every model under the same subject-level protocol. Of its 26 leaderboards we ran the eight scalp and two intracranial (HUP) boards that meet an inclusion criterion fixed in advance (Appendix B.1), seven of the benchmark's 21 datasets; every rank below is over these ten boards. Where a dataset offers more channels than pre-training saw, we choose the montage on the validation folds alone, never on test. On the scalp boards UniPopT has the best mean AUROC rank of all 18 models, 2.75 against 3.50 for the next best, REVE, and also the best mean rank over all three metrics (last block of Table 2). It is in the top three in most cells of the table, and on the two intracranial boards it has the best mean accuracy rank and the second-best mean AUROC rank, behind BrainWave.

5 Ablations and Analyses

Ablation design. In Sec. 4 we compared UniPopT with models from outside this paper; we now compare it with its own ablated variants, in both evaluation regimes. The pipeline has five replaceable parts: the encoder bank, the encoder, the positional encoding, the aggregator and its pre-training

Table 2: External evaluation on Brain4FMs (Shen et al., 2026). Percent, mean ± std over the benchmark's cross-validation folds; rank by value among the 18 models of the EEG leaderboards and the 11 models of the HUP leaderboards (17 / 10 published models plus ours). Cohen's κ is the benchmark's multi-class metric and is reported for the three-class Dep-STAI board only. Dep-122 rows use all 64 channels; UCSD rows use the 19 core channels. The last column is the best published model on the AUROC column. The bottom block gives the mean rank over the 8 EEG leaderboards, over the AUROC column and over every AUROC, Acc and F1 cell (24 cells), for UNIPOPT and, in the last column, for the best published model.

Board	AUROC	Rank	Acc	Rank	F1	Rank	κ	Rank	Best pub.
EEG leaderboards									
ADFD	**84.61** ± 4.72	**#1**	75.35 ± 2.18	#2	76.00 ± 4.55	#3			REVE 84.28
SD-28	81.35 ± 7.34	#2	74.48 ± 9.54	#4	76.75 ± 8.52	#4			BrainWave 88.16
MDD-64	**96.90** ± 2.47	**#1**	**89.47** ± 4.46	**#1**	89.95 ± 4.12	#2			REVE 96.88
UCSD-ON	**77.93** ± 6.40	**#1**	**70.37** ± 7.90	**#1**	**65.92** ± 21.59	**#1**			BrainWave 68.71
UCSD-OFF	61.44 ± 25.22	#4	**60.96** ± 19.63	**#1**	58.10 ± 28.59	#3			BrainOmni 63.57
Dep-BDI	64.70 ± 6.72	#5	63.15 ± 5.21	#4	47.89 ± 5.55	#6			BrainWave 72.40
Dep-STAI	**65.18** ± 3.41	**#1**	57.61 ± 3.83	#6	**43.75** ± 2.11	**#1**	**20.03** ± 6.56	**#1**	BrainWave 63.55
EEGMat	64.89 ± 7.12	#7	68.11 ± 12.45	#6	51.32 ± 7.92	#2			REVE 77.95
iEEG leaderboards									
HUP-ECoG	78.38 ± 9.09	#3	**81.12** ± 8.78	**#1**	65.80 ± 10.53	#2			BrainWave 87.93
HUP-SEEG	65.54 ± 4.27	#2	**75.25** ± 3.83	**#1**	40.09 ± 1.63	#2			BrainWave 78.57
Mean rank over the 8 EEG leaderboards (lower is better)									
AUROC	**2.75**								REVE 3.50
AUROC, Acc, F1	**2.88**								REVE 3.79

corpus. We vary three of them across the variants (Table 3) and keep the encoder architecture and the aggregator fixed. Each variant changes one part, except P4, which changes all three and is the original PopT recipe run on our data. We pre-train each variant once, at the same compute, and fine-tune each ablated variant with three seeds; we count a difference as zero when it does not exceed the larger of the two variants' standard deviations.

Table 3: The five pre-trained variants of the ablation. Each is a separate pre-training run of the same pipeline and differs from UNIPOPT only in the parts named here; the encoder architecture, the aggregator, the objectives, the optimiser and the list of scalp recordings are identical across variants. Details in Appendix A.

Variant	Encoder bank	Aggregator corpus	Position	What it isolates
UNIPOPT	Bimodal	Bimodal	Spherical	Full model
P1	Bimodal	EEG-only	Spherical	iEEG in the corpus
P2	EEG-only	EEG-only	Spherical	iEEG in the bank
P3	Bimodal	Bimodal	Cartesian	The positional encoding
P4	EEG-only	EEG-only	Cartesian	The original PopT recipe

In-domain ablation. On the official splits (Table 4) we can separate the two levels at which iEEG enters. Adding iEEG to the aggregator's pre-training corpus raises TUAB AUROC by +0.42 (UNIPOPT − P1), more than the seed standard deviation of either variant, whereas adding it to the encoder bank changes nothing (P1 − P2); on TUEV the corpus contrast is flat. The spherical encoding acts on the event task instead: against the Cartesian variant P3 we find TUAB within noise and TUEV κ higher by +2.04.

External ablation. We repeated the comparison under the same protocol on the nine Brain4FMs boards evaluated at the time (EEGMat was added later; Table 5, Fig. 2). UNIPOPT has the best mean AUROC rank of the five variants, and the aggregator corpus is again what matters: UNIPOPT − P1 is +2.62 AUROC, positive on 9 of 9 boards ($p = 0.001$), while the encoder bank again adds nothing. The spherical encoding raises F1 over P3 on 8 of 9 boards, but we find no significant AUROC gain ($p = 0.150$). The gain over the original PopT recipe (P4) is of the same size as the corpus contrast.

Where the gains come from. In both regimes, then, iEEG helps through the aggregator's pre-training corpus and not through the encoder's training bank: we see a small gain on the official splits,

Table 4: The five variants of Table 3 on the official TUAB and TUEV evaluation splits: AUROC (TUAB), weighted F1 and Cohen's κ (TUEV), in percent, mean $\pm$ std over fine-tuning seeds (5 for UniPopT, three per ablated variant). The lower block gives the differences between variants.

Variant	AUROC	Weighted F1	κ
UniPopT	89.52 $\pm$ 0.10	71.83 $\pm$ 0.26	62.39 $\pm$ 0.50
P1	89.10 $\pm$ 0.10	71.82 $\pm$ 0.60	62.12 $\pm$ 0.82
P2	89.01 $\pm$ 0.09	71.88 $\pm$ 1.31	62.42 $\pm$ 1.88
P3	89.20 $\pm$ 0.43	70.41 $\pm$ 1.41	60.36 $\pm$ 1.65
P4	88.88 $\pm$ 0.25	70.49 $\pm$ 1.22	60.66 $\pm$ 1.91
Aggregator corpus (UniPopT $-$ P1)	+0.42	+0.01	+0.27
Encoder bank (P1 $-$ P2)	+0.10	$-$0.06	$-$0.30
Positional encoding (UniPopT $-$ P3)	+0.32	+1.42	+2.04
All three (UniPopT $-$ P4)	+0.64	+1.34	+1.73

Table 5: The five variants of Table 3 on the nine Brain4FMs boards, AUROC in percent, mean $\pm$ std over folds, within-board rank of the five variants in parentheses.

Board	UniPopT	P1	P2	P3	P4
ADFD	84.61 $\pm$ 4.72 (2)	81.72 $\pm$ 5.91 (4)	85.38 $\pm$ 4.73 (1)	83.62 $\pm$ 4.24 (3)	79.72 $\pm$ 3.91 (5)
SD-28	81.35 $\pm$ 7.34 (1)	78.37 $\pm$ 8.95 (4)	80.48 $\pm$ 7.99 (2)	74.10 $\pm$ 18.51 (5)	79.02 $\pm$ 16.12 (3)
MDD-64	96.90 $\pm$ 2.47 (1)	95.35 $\pm$ 2.52 (3)	95.18 $\pm$ 3.88 (4)	96.18 $\pm$ 4.11 (2)	94.72 $\pm$ 3.06 (5)
UCSD-ON	77.93 $\pm$ 6.40 (1)	72.90 $\pm$ 10.15 (5)	73.62 $\pm$ 12.60 (4)	77.83 $\pm$ 3.83 (2)	74.05 $\pm$ 8.12 (3)
UCSD-OFF	61.44 $\pm$ 25.22 (2)	56.72 $\pm$ 18.87 (5)	60.28 $\pm$ 14.22 (3)	57.58 $\pm$ 13.55 (4)	61.67 $\pm$ 16.65 (1)
Dep-BDI	64.70 $\pm$ 6.72 (3)	61.71 $\pm$ 9.79 (5)	61.87 $\pm$ 6.28 (4)	69.13 $\pm$ 5.79 (1)	67.77 $\pm$ 4.52 (2)
Dep-STAI	65.18 $\pm$ 3.41 (1)	62.84 $\pm$ 1.85 (3)	61.22 $\pm$ 5.50 (5)	61.42 $\pm$ 3.83 (4)	63.71 $\pm$ 3.55 (2)
HUP-ECoG	78.38 $\pm$ 9.09 (2)	78.11 $\pm$ 7.90 (3)	77.62 $\pm$ 8.30 (4)	79.17 $\pm$ 9.14 (1)	75.82 $\pm$ 8.73 (5)
HUP-SEEG	65.54 $\pm$ 4.27 (1)	64.73 $\pm$ 5.04 (2)	61.63 $\pm$ 3.55 (3)	59.88 $\pm$ 5.02 (4)	55.97 $\pm$ 2.86 (5)
Nine-board mean	75.11	72.49	73.03	73.21	72.49
Mean rank	1.56	3.78	3.33	2.89	3.44
Friedman $\chi^2(4) = 10.84, p = 0.028$					

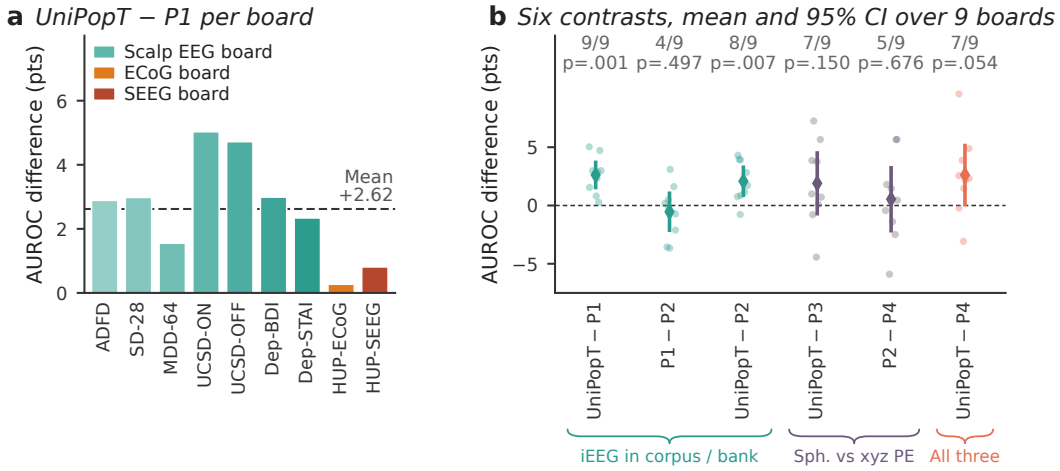


Figure 2: External decomposition on the nine Brain4FMs boards. In (b) dots are boards; above each contrast, the number of boards where it is positive and the paired p.

flat on TUEV, and about two and a half AUROC points on the external boards. BrainWave's modality ablation (Yuan et al., 2024) points the same way. Averaged this way, the spherical encoding looks modest: within noise on TUAB and on external AUROC, clearer on TUEV κ and external F1. One checkpoint nonetheless serves 19- to 64-channel scalp montages and 36- to 195-contact implants, and we examine below what the encoding does on positions that pre-training never saw. To see why the in-distribution gain of iEEG is small, we analysed the encoder's token space further.

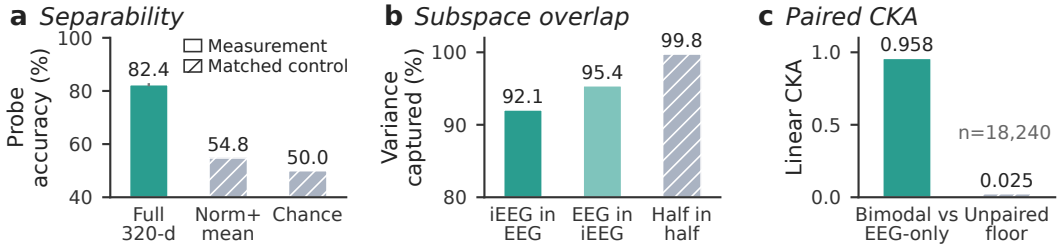


Figure 3: Geometry of the frozen bimodal encoder's token space (numbers in Table 19); hatched bars are the matched controls.

Redundancy in the encoder's token space. Two explanations are consistent with an in-distribution effect this small (+0.42 on TUAB, flat on TUEV; Table 4): the encoder may keep the modalities in separate subspaces that never interact ("two pipelines"), or it may embed iEEG into the same subspace that scalp EEG already fills ("redundancy"). We separate the two on the frozen bimodal encoder's per-channel tokens (Table 19, Fig. 3). A linear probe distinguishes EEG from iEEG tokens with 82.4% accuracy, far from the $\sim$100% we would expect from separate pipelines; given only each token's norm and mean, the same probe reaches 54.8%, so what separation exists lies in a token's direction, not in its scale or offset. We find 92.1% of the iEEG variance inside the top-20 principal subspace of the EEG tokens (the EEG-vs-EEG control gives 99.8%) and, decisively, a linear CKA (Kornblith et al., 2019) of 0.958 between the same scalp-EEG windows embedded by the bimodal and by the EEG-only encoder (unpaired floor 0.025). Training with iEEG did not change how the encoder represents scalp EEG.

We propose an explanation, without testing it here: the shared input space removes the content that distinguishes iEEG. Two preprocessing steps do this: 200 Hz resampling, which discards everything above 100 Hz and with it most of the 90–150 Hz high-gamma band, and per-window, per-channel z-scoring, which discards absolute amplitude. Both are field conventions rather than choices of ours: the z-score is TS2Vec's (Yue et al., 2022), and 200 Hz is the rate of LaBraM and BIOT (Jiang et al., 2024; Yang et al., 2023). The explanation is scoped to the size of the in-distribution effect: the external, corpus-level gain above is a property of the aggregator's pre-training corpus and is not addressed by an encoder-space measurement.

Table 6: UCSD-ON attribution of the 32-channel loss: UNIPOPT test AUROC in percent, same folds and learning-rate grid, crossing which channels are tokens with which serve as Laplacian references; last column, difference to the 19-channel base.

Tokens	Laplacian references	AUROC	vs 19-ch base
19 core	frozen standard-1005 table (pre-training montage)	77.93	+0.00
32	2 nearest of the 32-channel set (median 48 mm)	63.13	−14.80
19 core	2 nearest of the 32-channel set	66.76	−11.17
32	2 nearest at $\geq$ 50 mm	75.85	−2.08
19 core	2 nearest at $\geq$ 50 mm	77.19	−0.74
32	core 19 keep the frozen table, 13 extras 2 nearest	74.64	−3.29

Montage and recording length per subject. Two datasets offer more channels than the 19 of pre-training, and the validation folds choose opposite montages for them: all 64 channels on Dep-122, the 19 core channels on UCSD. The test folds agree: on Dep-122 the 64-channel input gains +7.43 AUROC on STAI and +6.36 on BDI, and on UCSD-ON the 32-channel input loses. The two datasets differ most in size, 122 subjects against 31. To see whether the value of the extra electrodes grows with the data, we fix the training subjects of Dep-STAI and vary how much of each recording they contribute: every subject keeps the first P % of its recording, from 0.4 to 8.17 minutes, with validation and test folds complete and the benchmark protocol unchanged (Fig. 4). With the spherical encoding the 64-channel input leads at every length and the gap opens with data: the 19-channel curve stays flat while the 64-channel curve keeps rising. With Cartesian coordinates (P3) the lead shrinks to about a quarter. Read the other way, the value of the encoding depends on the montage: UNIPOPT − P3 is +2.56 AUROC on 64 channels and +0.30 on the 19 channels of pre-training. The encoding

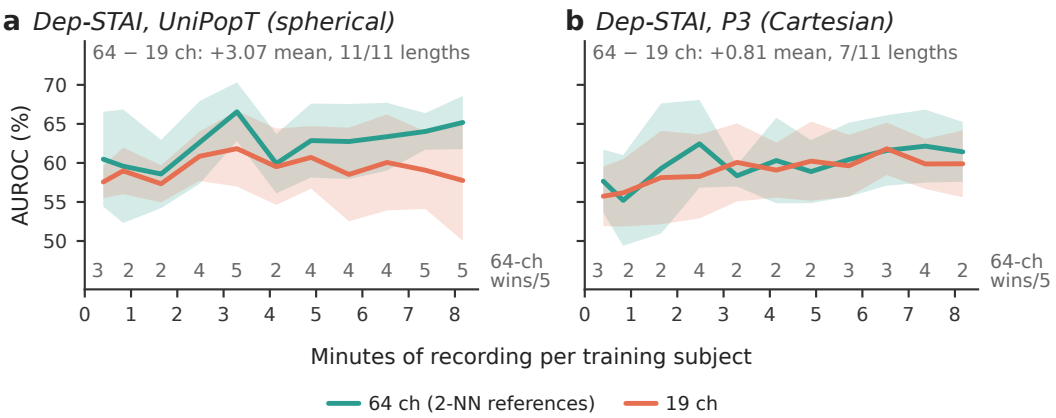


Figure 4: Dep-STAI AUROC against the minutes of recording per training subject, mean ± std over 5 folds.

is an interface: it changes little where every position is familiar and matters where most are not, which is also why it leaves TUAB unmoved. No published curve compares two montages along the fine-tuning data axis; the closest, AdaBrain-Bench (Wu et al., 2025), varies the number of subjects on one montage.

Reference geometry and montage transfer. On UCSD-ON, going from 19 to 32 channels lowers test AUROC by 14.80, and we can trace 11.17 of it to the Laplacian references alone: in the 32-channel set the two nearest neighbours sit closer than in pre-training (median 48 against 59 mm), and references taken at least 50 mm away recover most of the loss (Table 6). None of the work we reviewed varies the spacing of the Laplacian neighbours. The practical consequence is direct: whoever fine-tunes a pre-trained model on a denser montage should check the reference geometry first rather than assume that more channels are better.

6 Limitations and Future Work

- *Corpus diversity.* The bimodal corpus is also larger (4,043 against 3,076 recordings, at matched compute), so adding iEEG and adding recordings are confounded.
- *Head initialisation.* The benchmark's Kaiming fan-out head yields an initial logit std of about 20 on our unit-variance features, causing early stops; every competitor shares it.
- *Form of the iEEG gain.* Encoder-space diagnostics (Sec. 5) suggest that preprocessing shared with LaBraM and BIOT removes what is specific to iEEG, so the form of the gain is not attributed; an encoder trained at a sampling rate that keeps the high-gamma band, and one trained without the per-window, per-channel z-score, would test this and are left to future work. A scalp-heavy corpus with eight times the recordings (UniPopT-Ex, Appendix E.2) did not substitute for iEEG, lowering TUAB, TUEV and both intracranial boards and clearly raising only the two Dep-122 tasks, though it also changed the stride and the encoder.
- *One direction.* Only iEEG → scalp EEG is measured; the reverse lacks a matched control.

7 Conclusion

UniPopT reaches 89.52 ± 0.10 TUAB AUROC and the best mean AUROC rank on the scalp boards of Brain4FMs. Intracranial EEG helps it through the aggregator's pre-training corpus rather than the encoder bank, clearly on external boards and little in distribution, where iEEG tokens are largely redundant with scalp-EEG tokens; the spherical encoding pays off on electrode positions pre-training never saw, increasingly with the data recorded per subject.

AI USE STATEMENT

In this work, we used generative AI tools to implement methods (writing and editing the training, evaluation and analysis code, including the Brain4FMs wrapper and the table-generation script) and to clean and reformat data (consolidating the result files). We have not used generative AI tools to propose or refine hypotheses, to develop the conceptual framing, to design or give feedback on the methodology or experiments, to assist with translation, or to interpret results; formulating mathematical claims, proving them or writing proofs, generating synthetic data sets, and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to edit the paper for readability, to draft parts of the paper, and to identify and check related literature. All uses are also disclosed on the submission form. We have reviewed all AI-assisted work: generated code was run and its outputs checked against the raw result files; every number in the paper is produced from archived result files by one script the authors reviewed; AI-drafted text was revised by the authors; and every citation was checked against the cited work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Lindsay M. Alexander et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 4:170181, 2017.

Mehdi Azabou, Vinam Arora, Venkataramana Ganesh, Ximeng Mao, Santosh Nachimuthu, Michael Mendelson, Blake Richards, Matthew Perich, Guillaume Lajoie, and Eva L. Dyer. A unified, scalable framework for neural population decoding. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Donghong Cai, Junru Chen, Yang Yang, Teng Liu, and Yafeng Li. MBrain: A multi-channel self-supervised learning framework for brain signals. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2023.

Geeling Chau, Christopher Wang, Sabera Talukder, Vighnesh Subramaniam, Saraswati Soedarmadji, Yisong Yue, Boris Katz, and Andrei Barbu. Population transformer: Learning population-level representations of neural activity. In *International Conference on Learning Representations (ICLR)*, 2025.

Hsiang-Yun Sherry Chien, Hanlin Goh, Christopher M. Sandino, and Joseph Y. Cheng. MAEEG: Masked auto-encoder for EEG representation learning. In *NeurIPS Workshop on Learning from Time Series for Health*, 2022. arXiv:2211.02625.

Arnaud Delorme, Dung Truong, Choonhan Youn, Subhashini Sivagnanam, Claudia Stirm, Kenneth Yoshimoto, Russell A. Poldrack, Amit Majumdar, and Scott Makeig. NEMAR: An open access data, tools and compute resource operating on NeuroElectroMagnetic data. *Database*, 2022.

Berkay Döner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. LUNA: Efficient and topology-agnostic foundation model for EEG signal analysis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2510.22257.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A foundation model for EEG – adapting to any setup with large-scale pretraining on 25,000 subjects. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2510.21585.

Alexandre Gramfort et al. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7:267, 2013.

Danny Dongyeop Han, Yonghyeon Gwon, Ahhyun Lucy Lee, Taeyang Lee, Seong Jin Lee, Jubin Choi, Sebin Lee, Jihyun Bang, Seungju Lee, David Keetae Park, Shinjae Yoo, Chun Kee Chung, and Jiook Cha. DIVER-1: Scaling intracranial EEG foundation models for transferable representations. *arXiv preprint arXiv:2512.19097*, 2025.

Amir Harati, Meysam Golmohammadi, Silvia Lopez, Iyad Obeid, and Joseph Picone. Improved EEG event classification using differential energy. In *IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2015.

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.

Wei-Bang Jiang, Yansen Wang, Bao-Liang Lu, and Dongsheng Li. NeuroLM: A universal multi-task foundation model for bridging the gap between language and EEG signals. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2409.00101.

Kuntal Kokate, Bruno Aristimunha, Dung Truong, and Arnaud Delorme. Channel adaptation for EEG foundation models: A systematic benchmark across architectures, tasks, and training regimes. *arXiv preprint arXiv:2604.23091*, 2026.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International Conference on Machine Learning (ICML)*, 2019.

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

Gayal Kuruppu, Neeraj Wagh, Vaclav Kremen, Sandipan Pati, Gregory Worrell, and Yogatheesan Varatharajah. EEG foundation models: A critical review of current progress and future directions. *arXiv preprint arXiv:2507.11783*, 2025.

Dingkun Liu, Yuheng Chen, Zhu Chen, Zhenyao Cui, Yaozhi Wen, Jiayu An, Jingwei Luo, and Dongrui Wu. EEG-FM-Compass: Progress, benchmarking, and future directions for EEG foundation models. *arXiv preprint arXiv:2601.17883*, 2026.

Jingying Ma, Feng Wu, Qika Lin, Yucheng Xing, Chenyu Liu, Ziyu Jia, and Mengling Feng. CodeBrain: Bridging decoupled tokenizer and multi-scale architecture for EEG foundation model. In *International Conference on Learning Representations (ICLR)*, 2026. arXiv:2506.09110.

Iyad Obeid and Joseph Picone. The Temple University Hospital EEG data corpus. *Frontiers in Neuroscience*, 10:196, 2016.

Fanqi Shen, Enhong Yang, Jiahe Li, Junru Hong, Xiaoran Pan, Zhizhang Yuan, Meng Li, and Yang Yang. Brain4FMs: A benchmark of foundation models for electrical brain signal. *arXiv preprint arXiv:2602.11558*, 2026.

Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio Cases, and Andrei Barbu. BrainBERT: Self-supervised representation learning for intracranial recordings. In *International Conference on Learning Representations (ICLR)*, 2023.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. EEGPT: Pretrained transformer for universal and reliable representation of EEG signals. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

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

Jiamin Wu, Zichen Ren, Junyu Wang, Pengyu Zhu, Yonghao Song, Mianxin Liu, Qihao Zheng, Lei Bai, Wanli Ouyang, and Chunfeng Song. AdaBrain-Bench: Benchmarking brain foundation models for brain-computer interface applications. *arXiv preprint arXiv:2507.09882*, 2025.

Qinfan Xiao, Ziyun Cui, Chi Zhang, Siqi Chen, Wen Wu, Andrew Thwaites, Alexandra Woolgar, Bowen Zhou, and Chao Zhang. BrainOmni: A brain foundation model for unified EEG and MEG signals. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2505.18185.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Biosignal transformer for cross-data learning in the wild. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Joel Ye, Fabio Rizzoglio, Xuan Ma, Adam Smoulder, Hongwei Mao, Gary Blumenthal, William Hockeimer, Nicolas Kunigk, Dalton Moore, Patrick Marino, Raeed Chowdhury, J. Patrick Mayo, Aaron Batista, Steven Chase, Michael Boninger, Charles Greenspon, Andrew B. Schwartz, Nicholas Hatsopoulos, Lee Miller, Kristofer Bouchard, Jennifer Collinger, Leila Wehbe, and Robert Gaunt. A generalist intracortical motor decoder. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Ke Yi, Yansen Wang, Kan Ren, and Dongsheng Li. Learning topology-agnostic EEG representations with geometry-aware modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Yang You, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh. Large batch optimization for deep learning: Training BERT in 76 minutes. In *International Conference on Learning Representations (ICLR)*, 2020.

Zhizhang Yuan, Fanqi Shen, Meng Li, Yuguo Yu, Chenhao Tan, and Yang Yang. BrainWave: A brain signal foundation model for clinical applications. *arXiv preprint arXiv:2402.10251*, 2024. v7, September 2025; v1 (February 2024) was titled Brant-2: Foundation Model for Brain Signals.

Ayse Betul Yuce and Sebastian Stober. Benchmarking positional encoding strategies for transformer-based EEG foundation models. *arXiv preprint arXiv:2605.29754*, 2026.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. TS2Vec: Towards universal representation of time series. In *AAAI Conference on Artificial Intelligence*, 2022.

Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foundation model for intracranial neural signal. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Daoze Zhang, Zhizhang Yuan, Junru Chen, Kerui Chen, and Yang Yang. Brant-X: A unified physiological signal alignment framework. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024.

A DATA AND IMPLEMENTATION

The pre-training corpus and the encoder bank, and every implementation and evaluation detail behind Secs. 3 and 4.

A.1 PRE-TRAINING CORPUS

Table 7: Aggregator pre-training corpus. Hours are 5 s window counts converted to hours; TUH subject counts are not available in the manifest (1 = unknown). The last column is the number of recordings removed because they belong to the official TUAB / TUEV evaluation splits.

Dataset	Source	Modality	Recordings	Subjects	Hours	Eval removed
TUAB train	TUH	Scalp EEG	2,717	–	1,035.9	276
TUEV train	TUH	Scalp EEG	359	–	100.1	159
ds004752	NEMAR	SEEG + scalp	68	15	3.0	0
ds005415	NEMAR	iEEG	13	13	3.9	0
ds005522	NEMAR	iEEG	88	53	72.7	0
ds004789 (RAM)	NEMAR	iEEG	458	222	372.5	0
ds004809	NEMAR	iEEG	340	169	239.3	0
Total			4,043		1,827.4	

Aggregator corpus and leakage gate. Table 7 lists the aggregator corpus: the training portions of TUAB and TUEV (Obeid & Picone, 2016; Harati et al., 2015) and five public iEEG datasets from NEMAR (Delorme et al., 2022) (ds004752, ds004789, ds004809, ds005415, ds005522), 4,043 recordings and 1,827 h in total (1,136 h scalp EEG, 691 h iEEG). Every recording of the official TUAB and TUEV evaluation splits (435 recordings) is excluded at the manifest level, so no evaluation waveform is seen during self-supervised pre-training. The EEG-only corpus of the decomposition (variants P1, P2 and P4) is the same manifest with the iEEG rows removed: 3,076 scalp recordings against 4,043 for the bimodal corpus, the difference being the 967 iEEG recordings. Compute is matched across variants (300k steps, batch 256), so the bimodal variant samples from a larger and more diverse pool of recordings; this is the corpus-diversity confound of the corpus-level contrast (Sec. 5, Table 5), and the clean control, an equally large EEG-only corpus, was not run. Brain4FMs subjects are not in any pre-training set.

Encoder banks. Following the PopT protocol, which likewise trains TS2Vec on its own unlabelled data, the encoder is trained on a *bank* $\mathcal{B}$ of $N_{\mathcal{B}} \approx 200$k random 5 s single-channel (window, channel) segments drawn from the pre-training recordings of the same manifest: 40 per scalp recording and 80 per iEEG recording, the latter doubled so that the bank's EEG:iEEG ratio, $\approx$ 61:39, matches the window ratio of the corpus (200,400 segments for the bimodal bank); the EEG-only bank draws 65 per scalp recording (199,940 segments), so that both banks contain $\sim$200k segments. The bank composition (bimodal or EEG-only, at equal size) is one of the parts varied by the decomposition and the first level of the decomposition in Sec. 5.

UNIPOPT-EX. UNIPOPT-EX adds HBN (26,647 recordings, 2,094 h) and two further scalp corpora (1,157 recordings, 107 h) to the main corpus, less five iEEG recordings absent from the UNIPOPT-EX manifest (31,842 recordings in all), and upsamples the iEEG segments $5\times$ in its bank (Appendix E.2).

A.2 IMPLEMENTATION DETAILS

Signal chain. Both modalities follow the PopT preprocessing: resampling to 2048 Hz, notch filtering at the line frequency and its harmonics, Laplacian re-referencing against the two nearest electrodes of the same group (for intracranial channels the two adjacent contacts, $i-1$ and $i+1$, of the same shaft, grid or strip, PopT's original rule; for scalp EEG the two nearest positions of the template), and segmentation into windows of $w = 5$ s. Each (window, channel) pair is then resampled to $f_s = 200$ Hz and z-scored, so that $T = f_s w = 1000$ samples and both modalities enter scale-free; both steps are field conventions, the per-window z-score TS2Vec's own (Yue et al., 2022),

200 Hz LaBraM's native rate (Jiang et al., 2024) and BIOT's official TUAB rate (Yang et al., 2023). Per-channel modality (scalp, ECoG, SEEG) is read from the BIDS channel and electrode tables, never inferred from channel names, because many intracranial subjects mix SEEG and ECoG.

Encoder. E is TS2Vec (Yue et al., 2022): ten dilated convolutional blocks of hidden width 64, 0.6M parameters, output dimension $d_E = 320$, trained with the official hierarchical contrastive objective for 20k iterations at batch 128 on the bank of Appendix A.1 and then frozen; a window's token is the max-pool over time of the encoder output, the canonical TS2Vec classification representation. One consequence of the 200 Hz input is used in Sec. 5: everything above the 100 Hz Nyquist limit, which includes most of the 90–150 Hz high-gamma band, the band most specific to intracranial recordings, is invisible to the encoder.

Positional encoding. Each token $E(x_i^t)$ is projected by W to the aggregator width $d = 512$ and summed with $\mathrm{pos}(i)$ of Eq. equation 2, eight slots of 64 dimensions, before the layer normalisation of Eq. equation 1. Scalp positions come from a frozen standard-1005 template of MNE (Gramfort et al., 2013), intracranial ones from each subject's BIDS electrode table. The spherical coordinates are quantised to 1 mm and 1°, depth is $\max(0,\ 85\,\mathrm{mm} - r)$ for intracranial contacts and 0 for scalp electrodes, categorical codes are spread far apart in the sinusoidal table so that adjacent rows do not collapse scalp and ECoG onto the same embedding, and every index is clamped so that a negative coordinate cannot wrap around the table. Pre-training uses the 19 electrodes of the 10–20 system, with PopT's isotropic 5 mm coordinate jitter applied before the spherical features are computed and applied identically to the spherical and the Cartesian variants, so the two variants see the same geometry and differ only in how they consume it. Intracranial coordinates are transformed to MNI where the dataset provides a transform and taken as approximately MNI otherwise, a tolerance in keeping with that jitter and quantisation.

Pre-training objectives. The aggregator A, PopT's transformer encoder unchanged, maps [CLS] and H^t (Eq. equation 1) to $\mathbf{y}_{\mathrm{cls}}$ and $\{\mathbf{y}_i\}_{i\in S}$, and is pre-trained with PopT's two objectives (Chau et al., 2025) through linear heads. An input joins window t on a channel set S_A ($s = 0$) and window t' on a disjoint set S_B ($s = 1$) of one recording, t' being the next window with probability $\frac{1}{2}$ and a random one otherwise; next-segment prediction asks the CLS token whether t' follows t ($z \in \{0, 1\}$). For replaced-token detection each token is, with probability $\rho/2$ each, replaced by a random channel's window from a random time ($u_i = 1$) or kept as a control ($u_i = 0$), forming the set R:

$$\mathcal{L} = \mathrm{BCE}\big(\sigma(\mathbf{w}_N^\top \mathbf{y}_{\mathrm{cls}}),\, z\big) + \frac{1}{|R|}\sum_{i\in R} \mathrm{BCE}\big(\sigma(\mathbf{w}_C^\top \mathbf{y}_i),\, u_i\big). \tag{3}$$

Aggregator and optimisation. The aggregator is PopT's transformer unchanged (six pre-norm layers, eight heads, width $d = 512$, feed-forward width 2048, GELU) with PopT's two objectives (Chau et al., 2025), Eq. equation 3, and its replacement rate $\rho = 0.1$. Each pre-training example draws $n \in \{10, 20, \ldots, 100\}$ channels of one recording (all of them when it has fewer) and splits them evenly into S_A and S_B; the next window of the grid starts 0.2 s after the anchor; replacement windows come from a third random window of the same recording with its channels shuffled. Optimisation is LAMB (You et al., 2020) at learning rate 5×10^{-4}, batch 256, 300k steps with 2.5% warm-up, bf16 autocast, and the pre-training window starts thinned to one every 2 s (stride 10 on the 0.2 s window grid, which bounds the pre-load). We keep the checkpoint with the lowest validation loss; the final checkpoint gives the same downstream numbers within seed noise.

A.3 Downstream wrappers, montage rule and recording-length subsets

TUAB and TUEV. The aggregator and a linear head on the CLS token are fine-tuned end-to-end on 5 s windows of the official training split, with a validation subset split off by subject, and evaluated once on the official evaluation split. We report window-level AUROC for TUAB and weighted F1 and Cohen's κ for TUEV.

Brain4FMs. The model is inserted into the benchmark's code as one more foundation model: the benchmark's preprocessing, subject-level folds, learning-rate grid, two-layer MLP head, early stopping, class-balanced loss and argmax decision rule are used unchanged (Shen et al., 2026)

(Appendix B.1). Our wrapper, the interface code that inserts the model into the benchmark's framework and adds no parameters, runs the signal chain and the frozen encoder, builds the tokens under the montage rule below (template coordinates for scalp datasets, the subject's electrode table for HUP), pools the CLS output over the 5 s windows of a sample (a HUP segment is encoded whole), and fine-tunes the aggregator. Nothing on our side changes the frozen encoder, the pre-trained weights or the benchmark's protocol; the montage is chosen on the validation folds under the rule of Appendix B.2.

Montage rule. At fine-tuning time the rule is montage-invariant: every channel whose name resolves in the frozen standard-1005 template can become a token, and each token's Laplacian reference is the mean of the two nearest channels of the kept set, so 32- and 64-channel montages are handled without any change to the model. Which channel set a dataset is read at, all resolvable channels by default or the 19 channels of pre-training, is chosen on the validation folds (Appendix B.2): Dep-122 is reported at 64 channels and UCSD at 19, with both number sets in Appendix D.1 (Table 15).

Fine-tuning data per subject. For the recording-length experiment, every training subject is kept and each keeps only the first $P\,\%$ of its own recording (consecutive windows in time order; `POPT_B4F_SUBSET_UNIT=time`), with validation and test folds complete and the benchmark protocol unchanged; the design and results are in Appendix D.2 (Fig. 4, Table 17).

A.4 Evaluation setup details

Benchmarks. TUAB and TUEV are evaluated on their official evaluation splits, which no pre-training corpus in this paper has seen (Sec. 3, Appendix A.1). The literature reports TUAB on 10 s segments and TUEV on event-centred 5 s segments; we use 5 s sliding windows. On TUAB the BIOT, LaBraM-Base and retrained BIOT rows of Table 1 use 10 s windows and the other rows 5 s. Absolute numbers are therefore comparable only within a tier of Table 1, and we say so wherever tiers are juxtaposed. The Brain4FMs inclusion criterion, fixed before any result was seen, the ten boards and the benchmark's protocol are in Appendix B.1; the montage choice is in Appendix B.2.

Metrics. AUROC, accuracy (Acc) and F1 as defined by each benchmark: window-level AUROC on TUAB; weighted F1 and Cohen's κ on TUEV; the benchmark's AUROC, Acc and F1 columns on Brain4FMs, with κ added for the three-class Dep-STAI board because the benchmark reports it for multi-class tasks. On TUEV the weighted F1 (71.83 ± 0.26 against 70.20 for the retrained BIOT) is a metric dominated by the background class, which is why κ is the number stated in the main text. Decisions are argmax for every model including ours. The benchmark's fourth binary metric is tabulated in Table 9; we never selected on it and do not report it in the main text.

Baselines. Tier (A) of Table 1 copies published numbers (Yang et al., 2023; Jiang et al., 2024; Chau et al., 2025); each is pre-trained on its own corpus and evaluated under its own protocol, and the PopT row is on the authors' own 80/10/10 split rather than the official evaluation split. The BIOT row is BIOT's strongest pre-trained variant, which pre-trains on six EEG corpora including the TUAB and TUEV training sets. Tier (B) is BIOT retrained from scratch with its official code, recipe and budget (100 epochs) on exactly our TUAB-train and TUEV-train recordings, then fine-tuned with its official scripts; it separates the effect of the pre-training corpus from the architecture. A LaBraM retrain was not affordable within the submission budget; we use BIOT's drop from its published number as the proxy for the data effect. Tier (C) is our model. The strongest internal baseline is P4, the original PopT recipe (EEG-only encoder bank, Cartesian coordinates, EEG-only corpus) run on our corpus and split, which gives the official-split number that the PopT paper does not.

Seeds and statistics. We pre-train each variant once, by design; the comparison of interest is between variants. The final model is fine-tuned with 5 random seeds (Tables 1 and 4) and the ablated variants with three, so every difference in Table 4 is between 5 seeds and three. All stds are population stds (ddof = 0), on the percent scale. Differences between variants are computed from the unrounded means, so a difference can be 0.01 off the difference of the printed means. A difference between two variants is called zero when it does not exceed the larger of the two variants' stds.

What the decomposition holds fixed. Table 4 holds the encoder architecture, aggregator, objectives, optimiser, budget and the list of scalp recordings fixed and varies only the three factors named in Sec. 5. With 19 template electrodes every injective positional code is equivalent; the spherical one is the interface to intracranial coordinates and 64-channel montages (Appendix D.2), and everything else in Table 4 is fixed. UNIPOPT − P1 on TUAB (+0.42) is larger than either variant's seed std (0.10 and 0.10), and on TUEV it is inside them; the reading is in Sec. 5.

B BRAIN4FMS: PROTOCOL, MONTAGE CHOICE AND RANKS

The ten boards and the inclusion criterion, the benchmark's protocol, the mean ranks and per-board margins, and the montage choice made on the validation folds.

B.1 BRAIN4FMS: DATASETS, INCLUSION CRITERION, PROTOCOL AND MEAN RANKS

Inclusion criterion. Brain4FMs (Shen et al., 2026) lists 21 datasets. We ran every dataset that satisfied both of two conditions fixed before any result was seen: (i) publicly downloadable from our compute cluster or from a permitted mirror, and (ii) the aggregator is applicable, i.e. the channels map onto the frozen standard-1005 template or the dataset supplies electrode coordinates. Every dataset that met the criterion is reported; none was dropped for its score. A third exclusion follows from the encoder: its window is a fixed 5 s, and boards whose benchmark segments are shorter than 5 s (EEGMMIDB, 4 s; SEED-IV, 4 s; BCI-2a, 3 s; Chisco, 3.3 s) were not evaluated, because padding them would alter the input relative to the other 18 models. Datasets that fail (i): SEED-IV (licence application required), Cogitate-CF (registration required), DEAP and ADHD-Child (licence or account required) and ISRUC (host unreachable from the cluster). Datasets that fail (ii): SWEC (4.8 TB of long-term recordings, contacts without electrode coordinates), CHB-MIT and SleepEDF (bipolar montage) and ADHD-Adult (two channels, pairs differing per segment); single-channel datasets are excluded by (ii) as well.

The ten boards. Seven datasets give ten leaderboards. One line each (task, subjects, channels, the benchmark's sampling rate and segment length). Every board uses the benchmark's subject-level folds, five per dataset except EEGMat's four.

- ADFD: Alzheimer's / frontotemporal dementia vs control, 88 subjects, 19 channels (scalp), 250 Hz, 10 s segments.
- SD-28: schizophrenia vs control, 28 subjects, 19 channels (scalp), 250 Hz, 5 s segments.
- MDD-64: major depression vs control, 64 subjects, 19 channels (scalp), 256 Hz, 10 s segments; 181 of the 193 files are used because 12 files are missing from the figshare archive and cannot be obtained by anyone.
- UCSD-ON and UCSD-OFF: Parkinson's disease vs control, on and off medication, 31 subjects (15 patients, 16 controls, stratified by class across the five folds), 32-channel recordings at 256 Hz (downsampled from 512 Hz), 10 s segments; reported on the 19 core channels (Appendix D.1).
- Dep-BDI and Dep-STAI: depression (BDI, binary) and anxiety (STAI, three classes) from the same 122-subject, 64-channel dataset, 250 Hz, 10 s segments; all 64 template-resolvable channels are tokens (Appendix D.1).
- EEGMat: mental arithmetic vs rest, 36 subjects, 19 channels (scalp), 250 Hz, 10 s segments; four folds, the benchmark's group count for this dataset.
- HUP-ECoG: seizure detection, 20 subjects, cortical grid / strip contacts with recorded coordinates; 54–125 contacts per subject.
- HUP-SEEG: seizure detection, 37 subjects, depth contacts with recorded coordinates; 36–195 contacts per subject.

Both HUP boards are sampled at 512 Hz in 2 s segments. The 5 s minimum of the inclusion criterion binds only the scalp pathway, which cuts every sample into fixed 5 s windows and could meet a shorter segment only by padding it. The intracranial boards take a second pathway with no minimum length: each segment is resampled to 2048 Hz, notch-filtered and Laplacian re-referenced, brought to 200 Hz and z-scored, and the whole sequence passes through the frozen encoder once to give one token per contact. The segments are neither padded nor cropped.

Table 8: Mean rank on Brain4FMs, computed over the leaderboards we ran (8 EEG, 2 iEEG) with our model inserted into each table; lower is better. Models present on every scalp leaderboard, the eight best by scalp AUROC rank; – marks a model absent from the HUP leaderboards.

Model	EEG AUROC	EEG Acc	iEEG AUROC	iEEG Acc
UNIPOPT	2.75	3.12	2.50	1.00
REVE	3.50	2.88	–	–
BrainWave	3.88	7.00	1.00	2.50
BrainOmni	4.50	7.25	–	–
LaBraM	5.62	10.00	9.00	7.00
EEGPT	6.50	8.75	–	–
BrainBERT	8.25	8.88	3.50	4.50
CBraMod	9.25	6.38	6.00	8.50

The scalp boards have 17 published competitors (18 models including ours); the two HUP boards have 10 (11 including ours). Mean ranks in the main text are computed over these ten boards, not over the benchmark's 21 datasets.

Protocol. The benchmark's protocol is used unchanged: subject-level cross-validation (five folds, four on EEGMat), encoder learning rate searched over $\{10^{-4}, 5\times10^{-5}, 10^{-5}, 5\times10^{-6}, 10^{-6}\}$ with the head at ten times that rate, at most 50 epochs, early stopping on validation loss with patience 5, batch 128 and seed 1, the benchmark's two-layer MLP head, class-balanced loss weights, argmax decisions, and per-fold selection of the learning rate by the main validation metric (AUROC on the binary boards, Cohen's κ on Dep-STAI); per fold, the learning rate with the best validation score at the early-stopping epoch is tested. The same head, initialised by the benchmark's code, is used for every model on the leaderboard; its interaction with our unit-variance features is disclosed in Sec. 6. Our model enters through a parameter-free wrapper: frozen encoder, pre-trained aggregator, no change to the signal chain between datasets (Appendix A.3).

B.1.1 MEAN RANKS, PER-BOARD MARGINS AND THE COMPARISON WITH BRAINWAVE

Table 8 gives the mean ranks behind Sec. 4. Over the ten boards we ran, not the benchmark's 21 datasets, the eight-board scalp mean AUROC rank is 2.75, alone first among the 18 models (REVE 3.50, BrainWave 3.88), and the mean Acc rank is 3.12, second to REVE (2.88). On the two intracranial boards, again over the boards we ran, the mean Acc rank is 1.00 and the mean AUROC rank 2.50, behind BrainWave (1.00), whose HUP AUROCs we do not approach (BrainWave 87.93 on ECoG, BrainWave 78.57 on SEEG). The scalp result is carried by UCSD-ON (first in all three metrics, +9.22 AUROC over the next model), MDD-64 (first in AUROC by +0.02 and in Acc, second in F1 by -0.18), ADFD (first in AUROC by +0.33) and Dep-STAI (first in AUROC, F1 and κ, by +1.63 AUROC over BrainWave). SD-28 is second in AUROC (-6.81 to BrainWave), Dep-BDI fifth, and UCSD-OFF fourth in AUROC and first in Acc on a board whose fold std is about 20 points or more on every metric (19.6 to 28.6); EEGMat is seventh in AUROC (-13.06 to REVE) and second in F1. On ADFD (Acc #2, F1 #3) and Dep-STAI (Acc #6) the AUROC rank is not matched by accuracy; Sec. 5 records an observation on the benchmark head's initialisation, shared by every model on the leaderboard, that we did not correct.

Coordinates against data. BrainWave (Yuan et al., 2024) pre-trained jointly on scalp EEG and iEEG without electrode coordinates, on a corpus of about 40,000 hours against our 1,827 hours. With more than twenty times less data, our coordinate interface takes the higher mean AUROC rank on the scalp side of the benchmark (2.75 against 3.88, over the eight scalp boards among the ten we ran), while on the intracranial side BrainWave's AUROC clearly leads (1.00 against 2.50) and we lead in accuracy; we read this as coordinates traded for data, not as one model being stronger. BrainWave's reported gain comes from a joint corpus about twice the size of its EEG-only variant's, in few-shot and cross-domain settings, and it was not evaluated on TUAB or TUEV (Appendix C, Table 14). REVE (El Ouahidi et al., 2025) covers scalp EEG at (x, y, z, t) and the Population Transformer (PopT) (Chau et al., 2025) iEEG at integer xyz; each has coordinates, and each covers one modality.

B.2 BRAIN4FMS: THE PROTOCOL AND THE MONTAGE CHOICE MADE ON VALIDATION

Every board is evaluated under the benchmark's own protocol, unchanged: its preprocessing, subject-level folds, learning-rate grid, head, early stopping, class-balanced loss and argmax decision rule (Shen et al., 2026) (Appendix B.1). The encoder, the signal chain and the pre-trained aggregator are identical for every dataset and every board. One thing is chosen on our side, the montage, on the validation folds alone.

How many channels. Two datasets offer more channels than the 19 the model was pre-trained on, and the montage is chosen the same way, on validation AUROC alone, averaged over the five learning rates of the benchmark's grid. On Dep-122 the 64-channel input wins on both boards: 57.18 against 54.84 on Dep-BDI and 64.50 against 60.88 on Dep-STAI, with all five learning rates favouring 64 channels on both.

On UCSD the ordering reverses, and by a wider margin: the 19 core channels reach 64.06 against 57.61 for the full 32-channel montage on UCSD-ON and 54.76 against 48.24 on UCSD-OFF, so the denser montage loses 6.45 and 6.52 points on validation.

Dep-122 is therefore reported with all 64 channels and UCSD with the 19 core channels. What the two choices do on the test folds is in Appendix D.1.

F2. Table 9 lists the benchmark's F2 column for the eight binary boards. F2 was never a selection target and never entered the main table; it is printed so that every metric the benchmark reports is available.

Table 9: Brain4FMs F2 score of the eight binary boards (companion to Table 2). F2 was never a selection target and never entered the main table; it is listed here so that every metric the benchmark reports is available. Percent, mean $\pm$ std over folds, rank by value against the published leaderboard; last column: best published model on F2.

Leaderboard	F2	Rank	Best published (F2)
EEG leaderboards			
ADFD	74.25 ± 9.99	#5	NeuroGPT 85.77
SD-28	76.61 ± 8.28	#5	REVE 85.69
MDD-64	88.82 ± 4.38	#9	BIOT 93.11
UCSD-ON	69.21 ± 27.40	#2	BIOT 81.37
UCSD-OFF	61.21 ± 30.51	#4	SppEEGNet 71.24
Dep-BDI	46.69 ± 10.95	#8	LaBraM 70.65
EEGMat	57.21 ± 9.08	#7	REVE 66.00
iEEG leaderboards			
HUP-ECoG	59.98 ± 6.07	#3	BrainWave 65.08
HUP-SEEG	32.90 ± 2.36	#2	BrainWave 45.01

C EXTERNAL DECOMPOSITION ON THE NINE BRAIN4FMS BOARDS

Design. Sec. 5 gives the design and the reading in brief. We ran the five variants of Table 4 on the nine boards evaluated at the time (EEGMat, added later, is not part of the decomposition) under the main-table protocol (Dep-122 with 64 channels, the HUP boards in iEEG mode, the other boards on the 19 core channels; per-fold learning rate by validation; the benchmark grid only; test never used). All 45 cells completed 25 of 25 folds; every number is recomputed from the per-fold logs. Table 5 gives AUROC per board and variant, Table 10 the six factor contrasts (one paired difference per board, $n = 9$), and Fig. 2 the two-panel picture; Acc and F1 per board, the Acc and F1 contrasts, the training read-outs and the comparison with BrainWave follow below. In Fig. 2 each contrast's 95% confidence interval uses the t distribution with 8 degrees of freedom, so it excludes 0 exactly when the paired $p < 0.05$.

Reading. Six points. (1) UNIPOPT is the best of the five variants on all three metrics: mean AUROC rank 1.56 (nine-board mean 75.11), mean Acc rank 1.78, mean F1 rank 1.67. The Friedman

Table 10: Factor contrasts on the nine external boards (AUROC, percent; one paired difference A − B per board, n = 9). Mean, median, paired $t(8)$ and its p; Holm correction over the six contrasts, and over the two pre-registered level contrasts (UNIPOPT − P1, P1 − P2) only; number of boards where A is better and the two-sided sign-test p; range of the mean when any one board is left out; means over the seven scalp boards and the two iEEG boards; the same contrast on TUAB (Table 4). Last row: UNIPOPT − P1 without the two UCSD boards.

Contrast	Meaning	Mean	Med.	$t(8)$	p	H(6)	H(2)	Wins	Sign	LOO	EEG-7	iEEG-2	TUAB
UNIPOPT − P1	Corpus	+2.62	+2.89	+4.90	0.001	0.007	0.002	9/9	0.004	+2.32/+2.91	+3.21	+0.54	+0.42
P1 − P2	Bank	-0.54	-0.16	-0.71	0.497	0.994	0.497	4/9	1.000	-0.99/-0.15	-1.20	+1.79	+0.10
UNIPOPT − P2	Both	+2.08	+1.72	+3.56	0.007	0.037	–	8/9	0.039	+1.81/+2.44	+2.01	+2.33	+0.51
UNIPOPT − P3	Sph/xyz, bim.	+1.90	+0.99	+1.59	0.150	0.450	–	7/9	0.180	+1.23/+2.69	+1.75	+2.44	+0.32
P2 − P4	Sph/xyz, EEG	+0.54	+0.46	+0.43	0.676	0.994	–	5/9	1.000	-0.10/+1.34	-0.38	+3.73	+0.13
UNIPOPT − P4	All three	+2.62	+2.33	+2.26	0.054	0.216	–	7/9	0.180	+1.75/+3.33	+1.64	+6.07	+0.64
UNIPOPT − P1	Without the two UCSD boards (n = 7)	+1.98	+2.34	+4.69	0.003	–	–	7/7	–	–	–	–	–

Table 11: External decomposition, accuracy and F1 (companion to Table 5; same runs). Percent, mean ± std over folds, within-board rank of the five variants in parentheses.

Accuracy Board	UNIPOPT	P1	P2	P3	P4
ADFD	75.35 ± 2.18 (1)	71.78 ± 3.33 (5)	74.92 ± 4.15 (3)	74.99 ± 5.45 (2)	72.88 ± 4.31 (4)
SD-28	74.48 ± 9.54 (1)	73.97 ± 5.74 (2)	72.49 ± 8.89 (4)	69.48 ± 10.66 (5)	72.78 ± 9.30 (3)
MDD-64	89.47 ± 4.46 (2)	88.73 ± 5.08 (4)	90.27 ± 5.27 (1)	89.11 ± 4.89 (3)	87.97 ± 5.13 (5)
UCSD-ON	70.37 ± 7.90 (1)	63.72 ± 13.44 (2)	58.10 ± 11.45 (5)	58.58 ± 4.17 (4)	59.85 ± 8.27 (3)
UCSD-OFF	60.96 ± 19.63 (1)	53.90 ± 13.07 (3)	53.48 ± 12.20 (4)	48.75 ± 8.99 (5)	55.67 ± 9.68 (2)
Dep-BDI	63.15 ± 5.21 (4)	63.60 ± 2.06 (2)	61.67 ± 3.73 (5)	64.47 ± 2.92 (1)	63.50 ± 2.61 (3)
Dep-STAI	57.61 ± 3.83 (2)	55.98 ± 2.90 (4)	56.64 ± 6.05 (3)	53.34 ± 5.76 (5)	60.69 ± 3.30 (1)
HUP-ECoG	81.12 ± 8.78 (1)	73.29 ± 11.69 (4)	72.03 ± 9.87 (5)	74.70 ± 18.12 (3)	77.52 ± 10.15 (2)
HUP-SEEG	75.25 ± 3.83 (3)	77.53 ± 0.81 (1)	75.69 ± 3.68 (2)	73.86 ± 1.30 (4)	73.24 ± 1.31 (5)
Nine-board mean	71.97	69.17	68.37	67.48	69.34
Mean rank	1.78	3.00	3.56	3.56	3.11
Friedman $\chi^2(4) = 7.64$, $p = 0.106$					
F1 Board	UNIPOPT	P1	P2	P3	P4
ADFD	76.00 ± 4.55 (3)	74.65 ± 3.06 (5)	77.26 ± 4.60 (1)	76.09 ± 5.78 (2)	74.89 ± 6.21 (4)
SD-28	76.75 ± 8.52 (1)	75.24 ± 5.72 (2)	74.33 ± 9.21 (3)	69.07 ± 11.58 (5)	72.01 ± 11.02 (4)
MDD-64	89.95 ± 4.12 (2)	88.85 ± 5.15 (5)	90.69 ± 4.98 (1)	89.66 ± 4.60 (3)	89.00 ± 4.50 (4)
UCSD-ON	65.92 ± 21.59 (1)	63.40 ± 16.37 (2)	46.69 ± 29.89 (5)	55.80 ± 20.49 (3)	55.36 ± 15.46 (4)
UCSD-OFF	58.10 ± 28.59 (1)	57.10 ± 12.84 (2)	56.68 ± 11.16 (3)	54.16 ± 13.85 (4)	50.42 ± 23.41 (5)
Dep-BDI	47.89 ± 5.55 (2)	41.24 ± 12.59 (5)	43.57 ± 7.49 (4)	45.60 ± 16.16 (3)	51.66 ± 6.27 (1)
Dep-STAI	43.75 ± 2.11 (2)	44.28 ± 2.39 (1)	39.72 ± 3.02 (4)	39.28 ± 3.87 (5)	42.90 ± 4.02 (3)
HUP-ECoG	65.80 ± 10.53 (1)	57.63 ± 11.06 (4)	55.47 ± 10.18 (5)	62.31 ± 13.88 (2)	57.73 ± 12.34 (3)
HUP-SEEG	40.09 ± 1.63 (2)	38.17 ± 3.78 (4)	41.00 ± 5.81 (1)	38.94 ± 4.49 (3)	31.53 ± 8.90 (5)
Nine-board mean	62.69	60.06	58.38	58.99	58.39
Mean rank	1.67	3.33	3.00	3.33	3.67
Friedman $\chi^2(4) = 8.80$, $p = 0.066$					

test over the five variants is significant for AUROC ($\chi^2(4) = 10.84$, $p = 0.028$) and not for accuracy ($\chi^2(4) = 7.64$, $p = 0.106$) or F1 ($\chi^2(4) = 8.80$, $p = 0.066$). (2) The one factor that stands is iEEG in the aggregator corpus, UNIPOPT − P1: AUROC +2.62, 9/9 boards positive, paired $t(8) = +4.90$, $p = 0.001$; leaving any one board out keeps the mean between +2.32 and +2.91; Acc +2.81 and F1 +2.63 move the same way. Within the two pre-registered level contrasts (UNIPOPT − P1 and P1 − P2) the Holm-corrected p is 0.002 (F1 0.049); when all six contrasts are corrected together it is 0.007. (3) The gain sits on the scalp boards, not the intracranial ones: +3.21 over the seven EEG boards against +0.54 over the two iEEG boards. The two largest per-board differences are the two UCSD boards (+5.03, +4.72); without them the mean is +1.98, 7/7 positive, $p = 0.003$, so the effect does not rest on them. (4) The encoder-bank level is zero: P1 − P2 is -0.54 AUROC ($p = 0.497$), and Acc and F1 are not significant either. The two-level sum UNIPOPT − P2 (+2.08) therefore decomposes into a corpus term of +2.62 and a bank term of -0.54 that is noise. (5) The spherical encoding acts on the decision metrics, not on the ranking, and only in the bimodal setting: UNIPOPT − P3 is +1.90 AUROC ($p = 0.150$, not significant) but +3.70 F1, 8/9 boards, $p = 0.011$ (Holm over six 0.068), and +4.50 Acc ($p = 0.026$); in the EEG-only setting P2 − P4 is zero on all three metrics (AUROC +0.54, $p = 0.676$). An unchanged AUROC with moving argmax metrics points to the placement

Table 12: Factor contrasts on the nine external boards, accuracy and F1 (companion to Table 10; same columns, percent, n = 9 boards).

Contrast	Mean	Median	$t(8)$	p	Holm(6)	Holm(2)	A better	Sign p	Leave-one-out	EEG-7	iEEG-2
Accuracy											
UNIPOPT − P1	+2.81	+1.63	+2.31	0.050	0.248	0.099	7/9	0.180	+2.18/+3.44	+2.82	+2.78
P1 − P2	+0.80	+1.26	+0.97	0.362	0.664	0.362	6/9	0.508	+0.20/+1.29	+0.59	+1.55
UNIPOPT − P2	+3.61	+1.48	+2.28	0.052	0.248	–	7/9	0.180	+2.52/+4.16	+3.40	+4.33
UNIPOPT − P3	+4.50	+4.27	+2.74	0.026	0.153	–	8/9	0.039	+3.53/+5.22	+4.67	+3.90
P2 − P4	-0.98	-1.75	-1.03	0.332	0.664	–	3/9	0.508	-1.41/-0.42	-0.82	-1.52
UNIPOPT − P4	+2.63	+2.01	+2.09	0.070	0.248	–	7/9	0.180	+1.64/+3.34	+2.58	+2.80
F1											
UNIPOPT − P1	+2.63	+1.51	+2.77	0.024	0.122	0.049	8/9	0.039	+1.94/+3.03	+1.94	+5.04
P1 − P2	+1.68	+0.42	+0.82	0.436	0.873	0.436	5/9	1.000	-0.20/+2.25	+2.26	-0.33
UNIPOPT − P2	+4.32	+2.42	+1.94	0.088	0.263	–	6/9	0.508	+2.45/+5.01	+4.20	+4.71
UNIPOPT − P3	+3.70	+3.49	+3.27	0.011	0.068	–	8/9	0.039	+2.90/+4.18	+4.10	+2.32
P2 − P4	-0.01	+1.69	-0.00	0.996	0.996	–	5/9	1.000	-1.20/+1.07	-1.04	+3.61
UNIPOPT − P4	+4.31	+4.74	+2.71	0.027	0.122	–	8/9	0.039	+3.52/+5.32	+3.16	+8.32

Table 13: Training read-outs of the external decomposition: per board and variant, the number of folds (of 5) that stopped at the first evaluation / the mean early-stopping epoch. P4 shows two anomalies: on ADFD all five folds stop at the first evaluation and on HUP-SEEG the mean stop epoch is 36; P4 also selects the smallest learning rate (1e-6) in 8 of 45 folds (UNIPOPT: 0), so the HUP-SEEG part of UNIPOPT − P4 partly reflects P4 not training under this protocol.

Board	UNIPOPT	P1	P2	P3	P4
ADFD	2 / 1.8	3 / 0.6	2 / 0.6	3 / 1.2	5 / 0.0
SD-28	2 / 1.0	1 / 5.6	2 / 2.0	3 / 1.0	2 / 0.8
MDD-64	1 / 3.2	4 / 0.6	1 / 1.8	1 / 2.0	2 / 1.0
UCSD-ON	0 / 8.6	0 / 9.0	0 / 6.0	1 / 5.8	0 / 12.8
UCSD-OFF	1 / 5.8	1 / 4.6	0 / 5.8	0 / 5.8	1 / 1.6
Dep-BDI	2 / 1.0	4 / 0.2	2 / 1.2	4 / 1.2	1 / 1.0
Dep-STAI	2 / 2.6	3 / 0.4	2 / 0.6	2 / 3.8	3 / 2.6
HUP-ECoG	1 / 3.0	0 / 5.4	1 / 3.8	0 / 5.8	1 / 4.8
HUP-SEEG	1 / 3.0	0 / 8.8	0 / 6.6	0 / 9.2	0 / 36.0
First-evaluation stops, of 45 folds	12	16	10	14	15

of the decision boundary rather than the ordering of the scores. The official splits agree: the same contrast leaves TUAB AUROC within noise (+0.32) while moving TUEV κ by +2.04 and weighted F1 by +1.42, beyond both variants' stds (Table 4). This interaction between positional encoding and corpus modality was not a pre-registered contrast and is reported as exploratory. (6) Against the original PopT recipe, UNIPOPT − P4 is +2.62 AUROC ($p = 0.054$, not significant) and +4.31 F1 ($p = 0.027$, 8/9). The intracranial part (+6.07) comes mostly from HUP-SEEG (+9.57), where P4 trained abnormally under this protocol (mean stop epoch 36; 8 folds picked the smallest learning rate; all 5 ADFD folds stopped at the first evaluation; Table 13), so that cell partly reflects P4 not training rather than a representation difference.

Caveats. Two must accompany the reading. First, UNIPOPT − P1 carries a corpus-diversity confound: compute is matched (both variants 300k steps at batch 256), but UNIPOPT samples from 4,043 recordings and P1 from 3,076, so "iEEG added" and "31 % more distinct recordings" are not separated; the clean control, an equally large EEG-only corpus, was not run, and UNIPOPT-EX (Appendix E.2) is only circumstantial because it changed the stride and the encoder as well. Second, the two sides are not the same size: in distribution the same contrast is +0.42 on TUAB, above both variants' seed stds (0.10 and 0.10) but a fraction of the external effect, and flat on TUEV (+0.27 κ, +0.01 weighted F1; Table 4). The claim is therefore +3.21 AUROC on the seven external scalp boards, +2.62 over all nine boards, and a small fraction of that on the official splits, no more. The comparison with BrainWave's own joint-versus-EEG-only ablation (Yuan et al., 2024), which points the same way at about 1.9 times the relative size, is given with its four caveats at the end of this appendix (Table 14).

Accuracy and F1. Table 11 gives the per-board accuracy and F1. UNIPOPT has the highest nine-board mean and the best mean rank on both: accuracy 71.97 (mean rank 1.78) against 69.17 / 3.00 (P1), 68.37 / 3.56 (P2), 67.48 / 3.56 (P3), 69.34 / 3.11 (P4); F1 62.69 (1.67) against 60.06 / 3.33, 58.38 / 3.00, 58.99 / 3.33, 58.39 / 3.67. The Friedman statistics are given in point (1) above.

Accuracy and F1 contrasts. Table 12 repeats the six contrasts on accuracy and F1. The corpus-level contrast UNIPOPT − P1 goes the same way as on AUROC: accuracy +2.81 ($p = 0.050$, Holm over the two pre-registered level contrasts 0.099) and F1 +2.63 (8/9 boards, $p = 0.024$, Holm(2) 0.049). The bank-level contrast P1 − P2 is zero on both (accuracy +0.80, $p = 0.362$; F1 +1.68, $p = 0.436$). The PE contrast in the bimodal setting, UNIPOPT − P3, is the one place where accuracy and F1 move while AUROC does not (numbers and reading in point (5) above; sign test on F1 0.039); in the EEG-only setting the same PE contrast, P2 − P4, is zero on accuracy (-0.98) and F1 (-0.01) as well. Against the original PopT recipe, UNIPOPT − P4 is +2.63 accuracy ($p = 0.070$) and +4.31 F1 (8/9, $p = 0.027$); its iEEG-side F1 part (+8.32) is dominated by HUP-SEEG, see the training read-outs.

Training read-outs. Table 13 lists, per board and variant, the folds that stopped at the first evaluation and the mean stop epoch. First-evaluation stops out of 45 folds: UNIPOPT 12, P1 16, P2 10, P3 14, P4 15. Two anomalies sit in P4: on ADFD all 5 folds stop at the first evaluation, and on HUP-SEEG the mean stop epoch is 36; P4 also selects the smallest learning rate (10^{-6}) in 8 of 45 folds (UNIPOPT: 0). The HUP-SEEG part of UNIPOPT − P4 (AUROC +9.57) therefore partly reflects P4 not training under this protocol rather than a representation difference alone. The first-evaluation stops are the head-initialisation effect disclosed in Sec. 5; it acts on every variant.

Comparison with BrainWave's joint-pre-training experiment. BrainWave (Yuan et al., 2024) (Sec. 2.4 and Fig. 5 of that paper) splits its model into BrainWave-EEG and BrainWave-iEEG and reports the relative AUROC gain of the joint model. Table 14 copies their numbers. Their corpus is 40,907 h (EEG 35,675.5 h, iEEG 5,231 h); the single-modality variants see 1.74 B (EEG) and 1.42 B (iEEG) patches. One counter-example is reported: on Absence-16 → Atonic-5 the joint model's balanced accuracy 44.55 is below BrainWave-EEG's 47.23. Four caveats apply to any comparison with our numbers. (1) Their gains are relative gains averaged per experiment (their Fig. 5e), not absolute points; +7.25 % corresponds to roughly 5–6 AUROC points. (2) Their gains are measured in cross-domain and few-shot settings, where baselines are low. (3) The joint model sees about twice the data of the EEG variant (1.74 B + 1.42 B against 1.74 B patches), so modality and volume are mixed. (4) Their evaluation sets do not overlap with our nine boards, and TUAB / TUEV were not evaluated. Our corpus-level contrast UNIPOPT − P1 (frozen encoder, matched compute) is +2.62 absolute AUROC points, about +3.8% relative ($p = 0.001$), with the gain on the scalp boards (+3.21) rather than the iEEG boards (+0.54). The direction agrees with theirs, the size is about half, and the two-level design localises the effect to the aggregator-corpus level (bank level -0.54, $p = 0.497$), which their design cannot separate.

Table 14: Relative AUROC gains of BrainWave's joint model over its single-modality variants, as reported in BrainWave (Yuan et al., 2024) Sec. 2.4 / Fig. 5, for comparison with our corpus-level contrast (UNIPOPT − P1 = +2.62 absolute AUROC points on nine external boards). Their corpus: 40,907 h (EEG 35,675.5, iEEG 5,231); patches seen by the single-modality variants: 1.74 B (EEG) and 1.42 B (iEEG). See the four caveats in the text.

Joint model vs	Setting	Relative AUROC gain
BrainWave-EEG	Cross-domain	+7.25 %
BrainWave-iEEG	Cross-domain	+2.05 %
BrainWave-EEG	Few-shot (3 / 8-shot)	+8.30 %
BrainWave-iEEG	Few-shot	+4.98 %
BrainWave-EEG	Apnea-ECG (other recording type)	+19.48 %
BrainWave-iEEG	Apnea-ECG	+9.90 %
Counter-example: Absence-16 → Atonic-5, joint BACC 44.55 vs BrainWave-EEG 47.23		

D MONTAGE AND RECORDING LENGTH

One story in two number sets: what the montage does on the test folds, and how the two montages scale with the fine-tuning data recorded per subject.

D.1 MONTAGE EVIDENCE

The fine-tuning montage rule (Sec. 3; stated in full in Appendix A.3) lets a dataset be read at any channel set that resolves in the template. Which set is used is decided on the validation folds (Appendix B.2): all 64 channels on Dep-122, the 19 core channels on UCSD. This appendix gives what the two choices do on the test folds.

Dep-122, 19 vs 64 channels. Table 15 is the same model (UNIPOPT, same folds, same LR grid) fine-tuned on the 19 core channels and on all 64 template-resolvable channels. All 64 channels resolve. AUROC moves from 58.34 to 64.70 on Dep-BDI (+6.36, rank #10 $\rightarrow$ #5) and from 57.75 to 65.18 on Dep-STAI (+7.43, #5 $\rightarrow$ #1); accuracy, F1 and κ move in the same direction on both boards. This is the only comparison in the paper in which the same model is given more channels and nothing else changes; it shows that the 19-channel deficit on these boards was a channel loss, not a model limit.

Table 15: Dep-122 with 19 and with 64 channels: the same model (UNIPOPT, same folds, same LR grid) fine-tuned on the 19 core channels and on all 64 template-resolvable channels (2-nearest Laplacian references of the kept set). Percent, mean $\pm$ std over folds; rank by value against the published leaderboard. The 64-channel numbers are the main-table entries (Table 2).

Leaderboard	Metric	19 ch	Rank	64 ch	Rank	64 − 19
Dep-BDI	AUROC	58.34 ± 7.85	#10	64.70 ± 6.72	#5	+6.36
Dep-BDI	Acc	61.26 ± 2.43	#10	63.15 ± 5.21	#4	+1.89
Dep-BDI	F1	41.72 ± 11.37	#9	47.89 ± 5.55	#6	+6.17
Dep-STAI	AUROC	57.75 ± 7.74	#5	65.18 ± 3.41	#1	+7.43
Dep-STAI	Acc	49.07 ± 8.44	#11	57.61 ± 3.83	#6	+8.54
Dep-STAI	F1	38.02 ± 6.62	#5	43.75 ± 2.11	#1	+5.73
Dep-STAI	κ	10.48 ± 9.33	#3	20.03 ± 6.56	#1	+9.55

UCSD-ON, attribution of the 32-channel loss. On UCSD the same rule gives 32 tokens and a 32-channel base of 63.13 AUROC against 77.93 for the 19-channel base (-14.80), the same ordering as on the validation folds (Appendix B.2), so UCSD is reported with 19 channels. Table 6 crosses the two things the rule changes: which channels are tokens (19 core or all 32) and which channels serve as Laplacian references (the frozen pre-training table, the two nearest of the 32-channel set with median spacing 48 mm against 59 mm in pre-training, or the two nearest at ≥ 50 mm). Reading: keeping the 19 tokens but taking the near references of the 32-channel set (66.76) costs -11.17 on its own, so the reference spacing is the main cause; keeping the 32 tokens but moving the references to ≥ 50 mm recovers 75.85; the 13 new-position tokens, at positions never seen in pre-training, cost about -3.29 (with the core-19 inputs kept bit-identical to the 19-channel base); and 19 tokens with far references (77.19) return to the base value, within -0.74 of the 19-channel base, which is the consistency check.

D.2 RECORDING LENGTH PER SUBJECT AND THE 64-CHANNEL MONTAGE: FULL TABLES

Design. Sec. 5 gives the design and the reading in brief; this subsection gives every number. We fix the number of training subjects and vary the length of each recording: every training subject of every fold is kept and contributes the first P % of its recording (10 s segments in temporal order), $P = 5, 10, 20, \ldots, 90$; $P = 100$ is the main-table entry. Validation and test folds are complete. The benchmark protocol is unchanged: per-fold learning rate by validation κ at the early-stopping epoch on Dep-STAI and by validation AUROC on Dep-BDI, five folds per point. The prefix is deterministic (no sampling seed). The recording length per subject in minutes is given next to P in every table;

the full recording is 8.17 min per subject. None of these numbers enters the main table. Fig. 4 shows the Dep-STAI curves; Table 16 gives the per-length differences averaged over all lengths and over the short (0.4–1.7 min per subject), middle (2.5–4.1 min) and long (6.5–8.2 min) ranges; the per-length tables and the Dep-BDI curve follow. No published curve compares two montages along the fine-tuning data axis, and per-subject fine-tuning curves for foundation models are almost absent (Appendix F; AdaBrain-Bench (Wu et al., 2025) is the one we found).

Table 16: Recording-length curves, summary of the per-length differences (percent). Mean over the 11 lengths, number of lengths with a positive difference, and the mean over the short (0.4–1.7 min per subject), middle (2.5–4.1 min) and long (6.5–8.2 min) lengths. Dep-STAI UNIPOPT and P3 curves are in Table 17, Dep-BDI in Table 18. Protocol curves only (per-fold LR by validation).

Board, variant	Contrast	Metric	Mean	Positive	Short	Middle	Long
AUROC							
Dep-STAI, UNIPOPT	64 ch − 19 ch	AUROC	+3.07	11/11	+1.59	+2.30	+5.23
Dep-STAI, P3	64 ch − 19 ch	AUROC	+0.81	7/11	+0.71	+1.23	+1.20
Dep-STAI, 64 ch	UNIPOPT − P3	AUROC	+2.56	9/11	+2.16	+2.66	+2.47
Dep-STAI, 19 ch	UNIPOPT − P3	AUROC	+0.30	6/11	+1.28	+1.59	-1.56
Dep-BDI, UNIPOPT	64 ch − 19 ch	AUROC	+1.91	7/11	+4.25	-3.84	+5.67
Acc							
Dep-STAI, UNIPOPT	64 ch − 19 ch	Acc	+1.36	7/11	-1.51	+0.51	+5.87
Dep-STAI, P3	64 ch − 19 ch	Acc	+0.37	6/11	-0.03	-0.56	+1.18
Dep-STAI, 64 ch	UNIPOPT − P3	Acc	+1.64	6/11	+1.57	+3.28	+1.92
Dep-STAI, 19 ch	UNIPOPT − P3	Acc	+0.64	6/11	+3.04	+2.20	-2.78
Dep-BDI, UNIPOPT	64 ch − 19 ch	Acc	+2.22	10/11	+3.86	+1.21	+1.16
F1							
Dep-STAI, UNIPOPT	64 ch − 19 ch	F1	+2.15	10/11	-0.45	+3.50	+3.71
Dep-STAI, P3	64 ch − 19 ch	F1	-0.45	4/11	+0.63	-2.05	-0.47
Dep-STAI, 64 ch	UNIPOPT − P3	F1	+2.75	9/11	+1.28	+4.43	+2.77
Dep-STAI, 19 ch	UNIPOPT − P3	F1	+0.15	5/11	+2.36	-1.13	-1.41
Dep-BDI, UNIPOPT	64 ch − 19 ch	F1	+1.96	8/11	-0.56	+0.06	+5.61

Dep-STAI, UNIPOPT: the shape. Table 17 and Fig. 4 give the UNIPOPT curves with 64 and with 19 channels. The 64-channel input is ahead of the 19-channel input at all 11 lengths (mean +3.07 AUROC, 11/11), but the lead is within fold noise up to about 1.7 min per subject (+1.59 over the short range, 2–3 of 5 folds) and consistent from about 2.5–3.3 min on (4–5 of 5 folds at every length except 4.1 min, where it is 2 of 5), reaching +5.23 over the long range. The 19-channel curve is flat in recording length: it stays between 57.33 and 61.83 AUROC from 0.4 to 8.2 min (Spearman with length 0.15). The 64-channel curve rises: a short-range mean of 59.55 over 5–20 %, 63.38 over 70–90 % (5.7–7.4 min) and 65.18 at 100 % (Spearman 0.70). We do not claim a minute threshold; what stands is the shape, the 19-channel input tops out quickly and the 64-channel input keeps improving with recording length.

Dep-STAI, P3: the encoding is an interface. With P3 (Cartesian encoding) on the same data and protocol, 64 − 19 channels shrinks to +0.81 (7/11) and no longer widens with length (+0.71 short, +1.20 long). Read the other way, UNIPOPT − P3 is +2.56 on the 64-channel input (9/11; Acc +1.64, F1 +2.75) and +0.30 on the 19-channel input (6/11; Acc +0.64, F1 +0.15). The 19-channel montage is the one pre-training saw and both encodings handle it; the 45 additional positions were never seen as scalp tokens, and the spherical encoding makes their tokens usable where the Cartesian one does not. This is the most direct evidence in the paper that the positional encoding is an interface: what it buys is the use of positions that pre-training never saw, not a better binary ranking on the montage pre-training did see (Secs. 3 and 5). Limits: one board, five folds per point, adjacent points move by 3–5 AUROC, of the same size as the fold std; the prefix is the start of each recording, whose state may differ from later parts; the high-anxiety class has six training subjects.

Table 17: Dep-STAI recording-length curves (AUROC, percent, mean ± std over 5 folds). Every training subject is kept and contributes the first P % of its recording (minutes per subject in the second column); validation and test folds are complete; the benchmark protocol is unchanged (per-fold LR by validation κ at the early-stopping epoch); 100 % is the main-table entry. Left: UNIPOPT with 64 and with 19 channels and the number of folds in which 64 channels is ahead; right: the same for P3 (Cartesian PE). Not part of the main table.

		UNIPOPT (spherical)				P3 (Cartesian)			
P (%)	Minutes	64 ch	19 ch	Δ	Wins	64 ch	19 ch	Δ	Wins
5	0.40	60.48 ± 6.07	57.56 ± 2.07	+2.92	3/5	57.66 ± 4.04	55.73 ± 3.81	+1.93	3/5
10	0.83	59.58 ± 7.26	58.98 ± 2.99	+0.60	2/5	55.20 ± 5.81	56.17 ± 4.32	-0.97	2/5
20	1.65	58.58 ± 4.37	57.33 ± 2.38	+1.25	2/5	59.29 ± 8.33	58.12 ± 5.98	+1.17	2/5
30	2.48	62.61 ± 5.30	60.85 ± 3.16	+1.76	4/5	62.45 ± 5.62	58.27 ± 5.38	+4.18	4/5
40	3.29	66.55 ± 3.77	61.83 ± 4.86	+4.72	5/5	58.35 ± 1.37	60.07 ± 5.00	-1.72	2/5
50	4.14	59.92 ± 3.81	59.51 ± 4.89	+0.41	2/5	60.31 ± 5.48	59.08 ± 3.53	+1.23	2/5
60	4.89	62.86 ± 4.74	60.71 ± 4.01	+2.15	4/5	58.89 ± 4.04	60.23 ± 5.07	-1.34	2/5
70	5.71	62.74 ± 4.81	58.52 ± 6.00	+4.22	4/5	60.43 ± 4.73	59.63 ± 3.96	+0.80	3/5
80	6.53	63.36 ± 4.34	60.06 ± 6.15	+3.30	4/5	61.60 ± 4.51	61.81 ± 3.36	-0.21	3/5
90	7.36	64.03 ± 2.34	59.07 ± 4.96	+4.96	5/5	62.15 ± 4.67	59.87 ± 3.21	+2.28	4/5
100	8.17	65.18 ± 3.41	57.75 ± 7.74	+7.43	5/5	61.42 ± 3.83	59.89 ± 4.29	+1.53	2/5

Dep-BDI. Table 18 and Fig. 5 repeat the UNIPOPT curves on the binary board of the same dataset. 64 channels is ahead overall (AUROC +1.91, 7/11; accuracy +2.22, 10/11) and at the full length (+6.36).

Table 18: Dep-BDI recording-length curves, UNIPOPT with 64 and with 19 channels (same design as Table 17; per-fold LR by validation AUROC on this binary board). AUROC and accuracy in percent, mean ± std over 5 folds; "64 ahead" counts the folds in which 64 channels has the higher AUROC.

		AUROC				Acc		
P (%)	min/subj.	64 ch	19 ch	Δ	64 ahead	64 ch	19 ch	Δ
5	0.40	58.09 ± 11.37	52.91 ± 9.95	+5.18	4/5	61.05 ± 8.13	57.12 ± 4.20	+3.93
10	0.83	56.57 ± 7.27	54.58 ± 8.22	+1.99	3/5	58.51 ± 6.29	56.48 ± 6.99	+2.03
20	1.65	59.76 ± 8.70	54.17 ± 10.66	+5.59	4/5	59.46 ± 7.69	53.83 ± 8.11	+5.63
30	2.48	54.99 ± 12.11	57.38 ± 9.95	-2.39	1/5	59.38 ± 5.46	58.60 ± 5.16	+0.78
40	3.29	54.25 ± 8.52	57.11 ± 9.62	-2.86	0/5	58.78 ± 9.30	56.98 ± 4.02	+1.80
50	4.14	55.29 ± 12.75	61.57 ± 5.52	-6.28	2/5	60.47 ± 5.59	59.41 ± 2.52	+1.06
60	4.89	63.10 ± 8.00	59.03 ± 8.53	+4.07	4/5	63.62 ± 3.87	60.91 ± 1.16	+2.71
70	5.71	58.51 ± 7.27	59.86 ± 4.52	-1.35	2/5	62.37 ± 4.02	59.39 ± 4.37	+2.98
80	6.53	65.70 ± 6.31	57.94 ± 9.76	+7.76	4/5	62.10 ± 4.69	59.77 ± 5.39	+2.33
90	7.36	63.48 ± 2.38	60.59 ± 7.95	+2.89	3/5	60.17 ± 3.29	60.90 ± 5.06	-0.73
100	8.17	64.70 ± 6.72	58.34 ± 7.85	+6.36	5/5	63.15 ± 5.21	61.26 ± 2.43	+1.89

E TOKEN-SPACE GEOMETRY AND THE CORPUS-SCALING VARIANT

The table, sampling and probes behind the token-space measurement of Sec. 5, and the eightfold scalp-heavy corpus that was not adopted.

E.1 GEOMETRY OF THE ENCODER'S TOKEN SPACE: TABLE, SAMPLING AND PROBES

Sampling and probes. Samples: 30 recordings per dataset, 16 windows per recording, up to 24 channels per window, seed 0, giving 18,240 tokens per modality after balancing. Separability: 5-fold logistic regression on the 320-d tokens; the two-feature control uses only the vector norm and mean. Subspace overlap: PCA bases of dimension 20 fitted per modality; the control splits the EEG tokens into two random halves. Paired CKA (Kornblith et al., 2019): the same (recording, window, channel) EEG tokens embedded by the bimodal-bank encoder and by the EEG-only-bank encoder; the floor is CKA between unpaired samples.

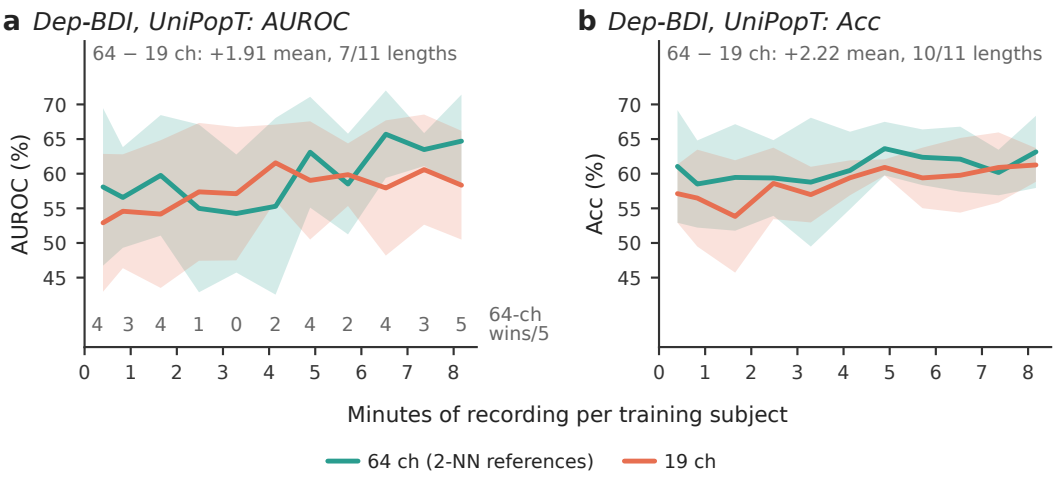


Figure 5: Dep-BDI recording-length curves, UniPopT with 64 and with 19 channels (AUROC, mean ± std over 5 folds, minutes per subject on the horizontal axis; protocol curves, per-fold LR by validation AUROC). Numbers in Table 18.

Table 19: Geometry of the frozen encoder's token space. Percent unless noted; separability is the accuracy of a 5-fold logistic regression, and captured variance is relative to the modality's own top-20 subspace. The right column is the matched control.

Test	Value	Control
EEG vs iEEG linear separability	82.4 ± 0.6	50 (chance)
Same probe, norm and mean only	54.8	50 (chance)
iEEG variance in the EEG top-20 PCA subspace	92.1	99.8 (EEG halves)
EEG variance in the iEEG top-20 PCA subspace	95.4	–
Median principal angle between the two subspaces (deg)	14.3	4.5 (EEG halves)
Centroid separation / within-modality spread	0.50	–
Linear CKA, same EEG windows, bimodal vs EEG-only encoder	0.958	0.025 (unpaired floor)

E.2 Scaling the corpus with scalp EEG (UniPopT-Ex)

With UniPopT-Ex we ask whether more data would move the numbers: the encoder bank grows from ∼200k to 555k segments (TUH, HBN (Alexander et al., 2017), two further scalp corpora, iEEG upsampled 5×; Appendix A.1) and the aggregator corpus from 4,043 to 31,842 recordings: the main corpus less five iEEG recordings absent from the UniPopT-Ex manifest, plus 26,647 HBN paediatric scalp recordings and 1,157 from two further scalp corpora. Table 20 shows the outcome under a decision rule fixed in advance (TUAB +0.5, or an improved Brain4FMs mean rank without an iEEG loss > 1): TUAB moves by -0.51, TUEV κ by -4.65, HUP-ECoG AUROC by -1.28 and HUP-SEEG AUROC by -5.47; only the two Dep-122 tasks, the scalp datasets with the most subjects, improve clearly (+6.10 / +3.10 AUROC, on the 19-channel input used at the time); UCSD-OFF moves by +5.58 inside a fold std above 20 points. The reading is that a scalp-dominated corpus pulls the shared representation toward scalp EEG at the expense of the intracranial tasks and of event-level TUEV, so more scalp data is not a substitute for the intracranial recordings. Two confounds are disclosed: UniPopT-Ex thins the pre-training window starts to one every 4 s instead of every 2 s (stride 20 instead of 10 on the 0.2 s grid; the stride-10 pre-load of 333 GB did not fit a 400 GB node) and retrains the encoder, so the loss cannot be attributed to the corpus alone. UniPopT-Ex was not adopted; every other number in this paper is UniPopT. The outcome is consistent with the saturating and sometimes non-monotonic scaling reported for EEG foundation models, reviewed in Appendix F.6.

Table 20: Scaling the pre-training corpus (UNIPOPT-EX) against UNIPOPT. UNIPOPT-EX retrains the frontend on a 555k-segment bank and pre-trains the aggregator on 31,842 recordings (UNIPOPT: ∼200k segments, 4,043 recordings), adding mostly scalp EEG (HBN, children). Top: official TUAB / TUEV eval splits, 5 fine-tuning seeds for UNIPOPT and three for UNIPOPT-EX. Bottom: Brain4FMs, same folds and learning-rate grid, rank against the benchmark leaderboard; the UNIPOPT column is the entry of Table 2, except on the two Dep-122 boards, where both sides use the 19-channel montage UNIPOPT-EX was run at. Percent; std over seeds / folds. Confounds: UNIPOPT-EX thins the pre-training window starts to one every 4 s (UNIPOPT: every 2 s) and retrains the encoder.

Task	Metric	UNIPOPT	UNIPOPT-EX	UNIPOPT-EX − UNIPOPT
TUAB	AUROC	89.52 ± 0.10	89.01 ± 0.06	-0.51
TUEV	Weighted F1	71.83 ± 0.26	67.84 ± 2.26	-3.99
TUEV	κ	62.39 ± 0.50	57.74 ± 2.23	-4.65

Dataset	Metric	UNIPOPT	Rank	UNIPOPT-EX	Rank	UNIPOPT-EX − UNIPOPT
ADFD	AUROC	84.61 ± 4.72	#1	81.59 ± 5.92	#3	-3.02
ADFD	Acc	75.35 ± 2.18	#2	72.70 ± 6.92	#3	-2.65
ADFD	F1	76.00 ± 4.55	#3	74.24 ± 4.67	#3	-1.76
SD-28	AUROC	81.35 ± 7.34	#2	75.45 ± 11.28	#4	-5.90
SD-28	Acc	74.48 ± 9.54	#4	72.34 ± 7.70	#4	-2.14
SD-28	F1	76.75 ± 8.52	#4	74.07 ± 7.71	#4	-2.68
MDD-64	AUROC	96.90 ± 2.47	#1	96.44 ± 2.38	#2	-0.46
MDD-64	Acc	89.47 ± 4.46	#1	90.22 ± 4.59	#1	+0.75
MDD-64	F1	89.95 ± 4.12	#2	90.89 ± 4.23	#1	+0.94
UCSD-ON	AUROC	77.93 ± 6.40	#1	74.70 ± 5.16	#1	-3.23
UCSD-ON	Acc	70.37 ± 7.90	#1	65.94 ± 6.16	#1	-4.43
UCSD-ON	F1	65.92 ± 21.59	#1	60.92 ± 18.13	#2	-5.00
UCSD-OFF	AUROC	61.44 ± 25.22	#4	67.02 ± 21.21	#1	+5.58
UCSD-OFF	Acc	60.96 ± 19.63	#1	62.11 ± 13.66	#1	+1.15
UCSD-OFF	F1	58.10 ± 28.59	#3	62.84 ± 15.73	#1	+4.74
HUP-SEEG	AUROC	65.54 ± 4.27	#2	60.07 ± 3.93	#4	-5.47
HUP-SEEG	Acc	75.25 ± 3.83	#1	74.90 ± 4.88	#1	-0.35
HUP-SEEG	F1	40.09 ± 1.63	#2	42.99 ± 3.70	#2	+2.90
HUP-ECoG	AUROC	78.38 ± 9.09	#3	77.10 ± 4.86	#3	-1.28
HUP-ECoG	Acc	81.12 ± 8.78	#1	78.58 ± 8.26	#2	-2.54
HUP-ECoG	F1	65.80 ± 10.53	#2	60.50 ± 10.39	#3	-5.30
Dep-BDI	AUROC	58.34 ± 7.85	#10	64.44 ± 4.91	#5	+6.10
Dep-BDI	Acc	61.26 ± 2.43	#10	58.42 ± 3.30	#14	-2.84
Dep-BDI	F1	41.72 ± 11.37	#9	49.47 ± 9.68	#5	+7.75
Dep-STAI	AUROC	57.75 ± 7.74	#5	60.85 ± 4.96	#2	+3.10
Dep-STAI	Acc	49.07 ± 8.44	#11	51.04 ± 5.33	#10	+1.97

F EXTENDED RELATED WORK

This appendix carries the full versions of the paragraphs condensed in Sec. 2.

F.1 SCALP-EEG FOUNDATION MODELS

Self-supervised scalp-EEG models began on fixed montages: BENDR (Kostas et al., 2021) adapts the contrastive wav2vec recipe to raw EEG, and MAEEG (Chien et al., 2022) reconstructs masked EEG features; both assume one channel set. MMM (Yi et al., 2023) maps heterogeneous montages onto a unified topology with geometry-aware position encoding, the first step toward the topology-agnostic designs of LUNA and REVE below. BIOT (Yang et al., 2023) tokenises each channel's spectrogram patches and pre-trains a linear-attention transformer across heterogeneous montages. LaBraM (Jiang et al., 2024) learns a neural tokenizer and masks tokens over roughly 2,500 hours of EEG. CBraMod (Wang et al., 2025) and EEGPT (Wang et al., 2024) refine the masked-modelling recipe with criss-cross attention and spatio-temporal alignment. NeuroLM (Jiang et al., 2025) couples an EEG tokenizer to a language model for multi-task instruction tuning. REVE (El Ouahidi et al., 2025) scales to the largest public EEG corpus with a four-dimensional (x, y, z, t) positional encoding, and LUNA (Döner et al., 2025) builds topology-agnostic representations. CodeBrain (Ma et al., 2026)

and BrainOmni (Xiao et al., 2025) are further entries on the external leaderboard we evaluate against (Sec. 4). All of these models are scalp-only, and their published TUAB / TUEV numbers come from corpora of different size and preprocessing (LaBraM alone pre-trains on about 2.2× our scalp-EEG hours). This is why Table 1 carries a data-matched retrained baseline (Sec. 4).

F.2 Intracranial models

BrainBERT (Wang et al., 2023) pre-trains a single-channel spectrogram encoder on SEEG. Brant (Zhang et al., 2023) models multi-channel iEEG with time- and channel-wise attention. MBrain (Cai et al., 2023) is a multi-channel self-supervised framework applied to both EEG and SEEG, but it trains one model per modality; it does not pre-train across the two. DIVER-1 (Han et al., 2025) fits data-limited scaling laws on iEEG and is discussed in Appendix F.6. Treating every sensor as a token that carries its own identity, so that a model can consume arbitrary sensor sets, was established for spiking populations by POYO (Azabou et al., 2023), which tokenises spikes with unit embeddings, and scaled by NDT3 (Ye et al., 2025); PopT brought the idea to intracranial field potentials. The Population Transformer (PopT) (Chau et al., 2025) keeps a frozen single-channel encoder and learns to aggregate an arbitrary set of electrodes, identified by their Cartesian coordinates, with a next-segment and a replaced-token objective. PopT reports an EEG variant on TUAB, but on the authors' own 80/10/10 split rather than the official evaluation split (Table 1, tier A).

F.3 Components we build on

We take PopT as the aggregator because its set-of-electrodes interface is what a bimodal model needs, and we keep its next-segment and replaced-token objectives. P4 in Table 4 is the original PopT recipe, including its integer Cartesian table-lookup positional encoding. TS2Vec (Yue et al., 2022) is our single-channel encoder, with its own per-window z-score kept (field conventions; Appendix A.2). BrainBERT is PopT's original intracranial encoder. BIOT is the data-matched retrained baseline of Table 1, and its 200 Hz rate is the rate of our pipeline (field conventions; Appendix A.2).

F.4 Joint EEG–iEEG pre-training: the BrainWave precedent

BrainWave (Yuan et al., 2024)[1] pre-trains one channel-independent model on EEG and iEEG and, in its Section 2.4, splits it into BrainWave-EEG and BrainWave-iEEG. The joint model gains a relative AUROC of +7.25 % over the EEG-only variant in cross-domain evaluation, with further gains in few-shot and cross-recording-type settings. Its corpus is 40,907 hours (35,675.5 EEG, 5,231 iEEG), against 1,827 hours here. The model has no electrode coordinates. The four caveats that apply when this precedent is read against our result, and the comparison with our corpus-level contrast, are in Appendix C (Table 14). Brant-X (Zhang et al., 2024) aligns EEG with simultaneously recorded peripheral signals through a two-level contrastive objective; it is the closest precedent for paired cross-modal alignment, but not for iEEG. Our decomposition holds the data budget fixed at both levels and uses coordinates (Sec. 5).

Two-modality models with a sensor interface exist outside iEEG: BrainOmni (Xiao et al., 2025) unifies EEG and MEG with a sensor encoder that carries spatial layout, orientation and sensor type, which is the closest precedent for our modality code; it contains no intracranial data and does not ask whether one modality helps the other.

F.5 Spatial encodings

Coordinate-based positional encodings are now common. REVE uses (x, y, z, t), PopT uses an integer Cartesian lookup, and LUNA is topology-agnostic by construction but reports sensitivity to an unseen montage. A recent study evaluates spherical coordinates on CBraMod and finds task-dependent effects without a consistent advantage (Yuce & Stober, 2026). Our spherical encoding adds depth and a modality code so that a scalp electrode and a depth contact at the same angular position are distinguishable (Sec. 3). Consistent with the CBraMod study, its effect is task-dependent and we treat it as an interface first: on the official splits it leaves the binary TUAB ranking within seed noise (+0.32) but moves the event task beyond it (TUEV κ +2.04, weighted F1 +1.42; Table 4),

[1] Earlier versions of the same manuscript (arXiv 2402.10251) were titled Brant-2.

and on the external boards UNIPOPT − P3 is +1.90 AUROC ($p = 0.150$; Table 10). Where the fine-tuning montage differs from pre-training the picture changes: on the 64-channel Dep-STAI input UNIPOPT − P3 is +2.56 AUROC against +0.30 on 19 channels (Table 17).

F.6 SCALING AND DATA EFFICIENCY

Pre-training scale saturates early or is non-monotonic in every EEG foundation-model paper that reports a curve. LaBraM-Base with 500 hours exceeds the 2,500-hour model on TUAB, and its Base curve is non-monotonic (about 0.819 at 500 hours, 0.796 at 1,000 and 0.814 at 2,500) (Jiang et al., 2024). CBraMod gains about 0.04 kappa on FACED over four decades of pre-training hours, with overlapping error bands (Wang et al., 2025). EEGPT fits power-law exponents near zero (Wang et al., 2024). NeuroLM's larger variants are worse on HMC, and all three variants lie below LaBraM-Base on TUAB, with instruction tuning as a confound (Jiang et al., 2025). PopT's pre-training-data curve flattens between 25 and 50 % of the corpus (Chau et al., 2025). Outside EEG, NDT3 reports negative scaling to 2,000 hours of intracortical data, attributed to interference from dissimilar recordings (Ye et al., 2025), and DIVER-1 finds that an in-domain iEEG corpus beats an out-of-domain corpus of the same size (AUROC 0.770 against 0.710), that a 15.8× larger corpus reaches 0.840 and then plateaus, and that subject count at a fixed budget is non-monotonic (Han et al., 2025). A critical review concludes that larger pre-training sets did not necessarily improve downstream performance (Kuruppu et al., 2025), and EEG-FM-Compass concludes that current EEG foundation models do not exhibit scaling-law behaviour (Liu et al., 2026). Our UNIPOPT-EX variant, a larger and scalp-heavy corpus that scores lower (Appendix E.2), is consistent with this literature, with the three simultaneous changes disclosed there. Fine-tuning data curves are usually drawn by sample, trial or k-shot count: PopT, BrainBERT, Brant, BrainWave, CBraMod, NeuroLM, REVE and Brain4FMs all sample this way. Curves drawn by number of training subjects for a fine-tuned foundation model exist only in AdaBrain-Bench (Wu et al., 2025), where LaBraM and CBraMod on sleep staging saturate after about 40 subjects and CBraMod on SHHS rises from 40.70 % with one subject to 63.79 % with 40. No published curve compares two channel sets or montages along the data axis. Our recording-length curves do (Fig. 4, Table 17, Appendix D.2).

F.7 MONTAGE ADAPTATION

Pre-training on few channels and fine-tuning on many has been measured twice. MMM (Yi et al., 2023) transfers from 32 to 62 channels with small gains and one negative transfer; BIOT transfers from 8 to 16 channels and is mostly worse, once below training from scratch (Yang et al., 2023). CBraMod pre-trains on 19 channels and fine-tunes on larger montages, as we do, but reports no 19-channel control (Wang et al., 2025). REVE reports a channel-count curve from 64 channels down to 1 at full data (El Ouahidi et al., 2025). A dedicated study compares four channel-adaptation methods across five models and five tasks, without varying data size (Kokate et al., 2026). No work crosses montage with fine-tuning data size. The 19- against 64-channel comparison on Dep-122 (Table 15) and the recording-length curves at both channel counts (Fig. 4) fill this gap (Appendix D.1).

F.8 EVALUATION PRACTICE

Brain4FMs (Shen et al., 2026) evaluates 17 foundation models on 21 public datasets under one subject-level cross-validation protocol with a fixed head and per-fold learning-rate selection on the validation fold. This removes the split and preprocessing differences that make published TUAB / TUEV numbers hard to compare. We adopt its protocol unchanged for the external evaluation, including the per-fold selection (Appendix B.1).