Learning to Dream in EEG

World models (LeJEPA) for biosignals

Thorir Mar Ingolfsson

Postdoctoral Researcher · ETH Zürich · Integrated Systems Laboratory

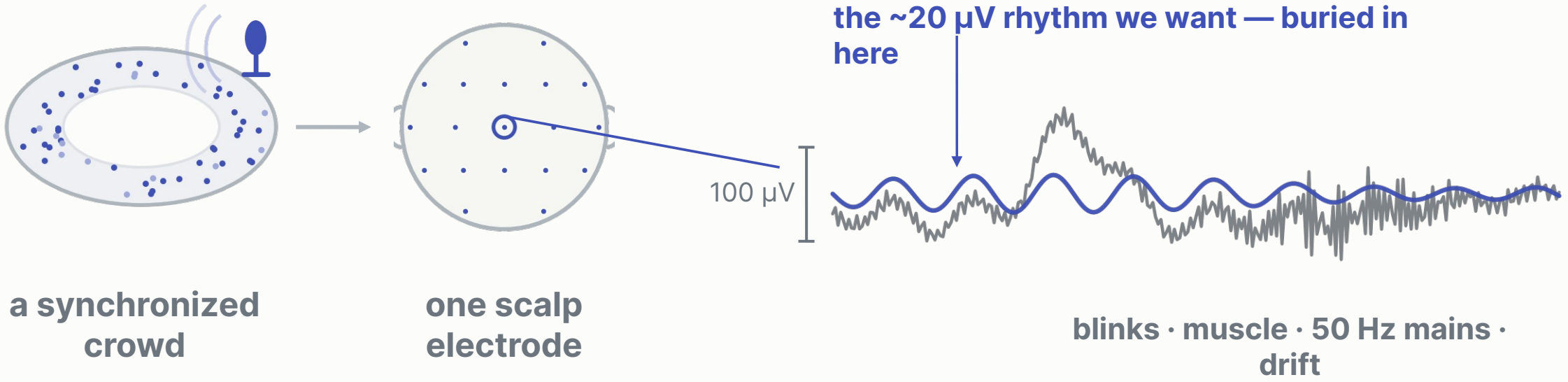
Forgis Research Night · DARE Campus, Schlieren · 24 June 2026

To predict what's hidden, a model must understand the signal.

This is the learning principle behind LeCun's world models.

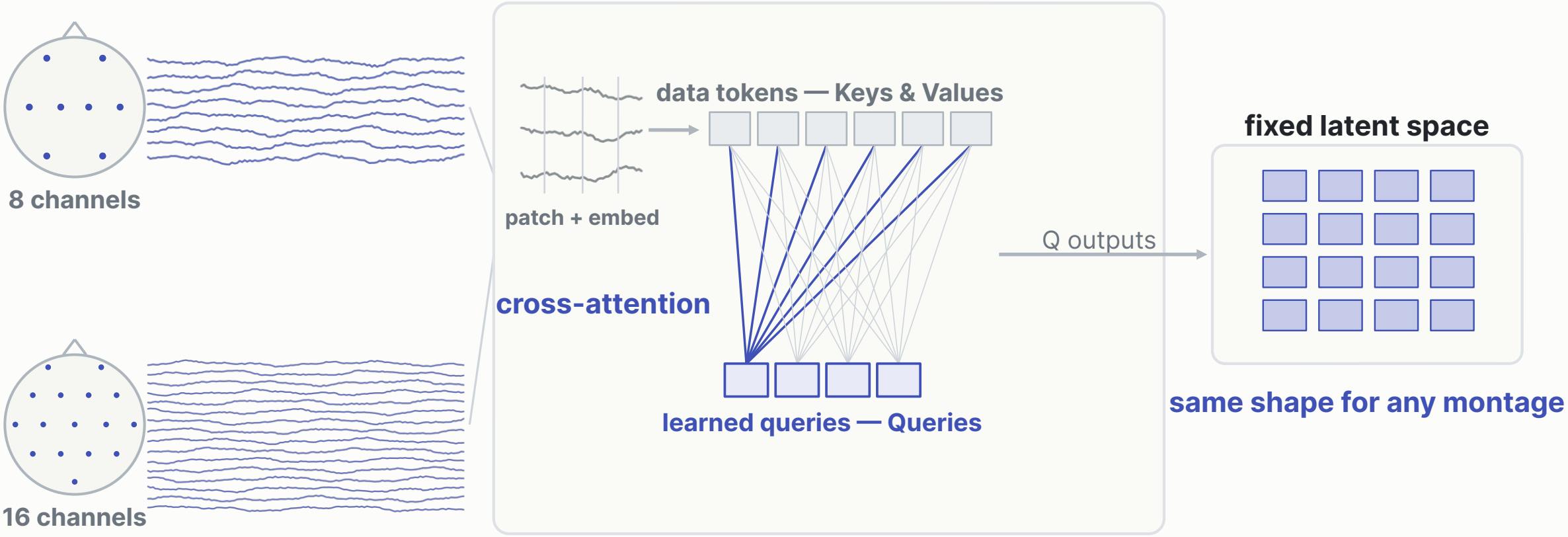
A full world model also predicts the future from actions, to plan. Biosignals are recorded, not acted on, so we borrow the half that learns structure.

EEG is the muffled roar of a neural crowd, heard through the skull.



The signal we want is buried in the recording: SNR < 1.

Channel unification: learned queries read any montage.

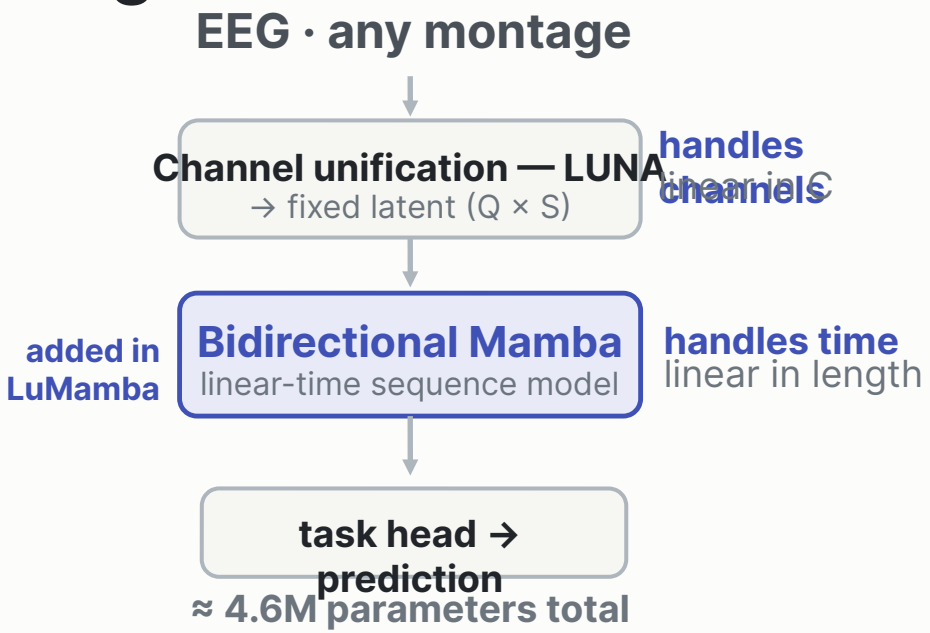


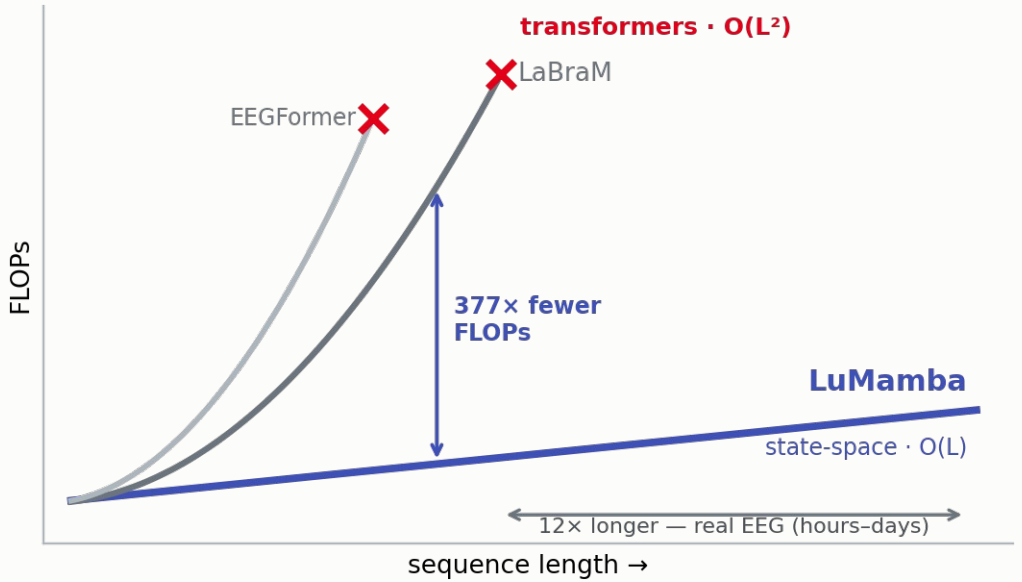
LuMamba reuses this: one model reads any EEG montage on Earth.

Cross-attention is linear in the number of channels: O(C·Q), never quadratic.

LUNA · Döner et al., NeurIPS 2025

LuMamba: channel unification for any montage, Mamba for any length.

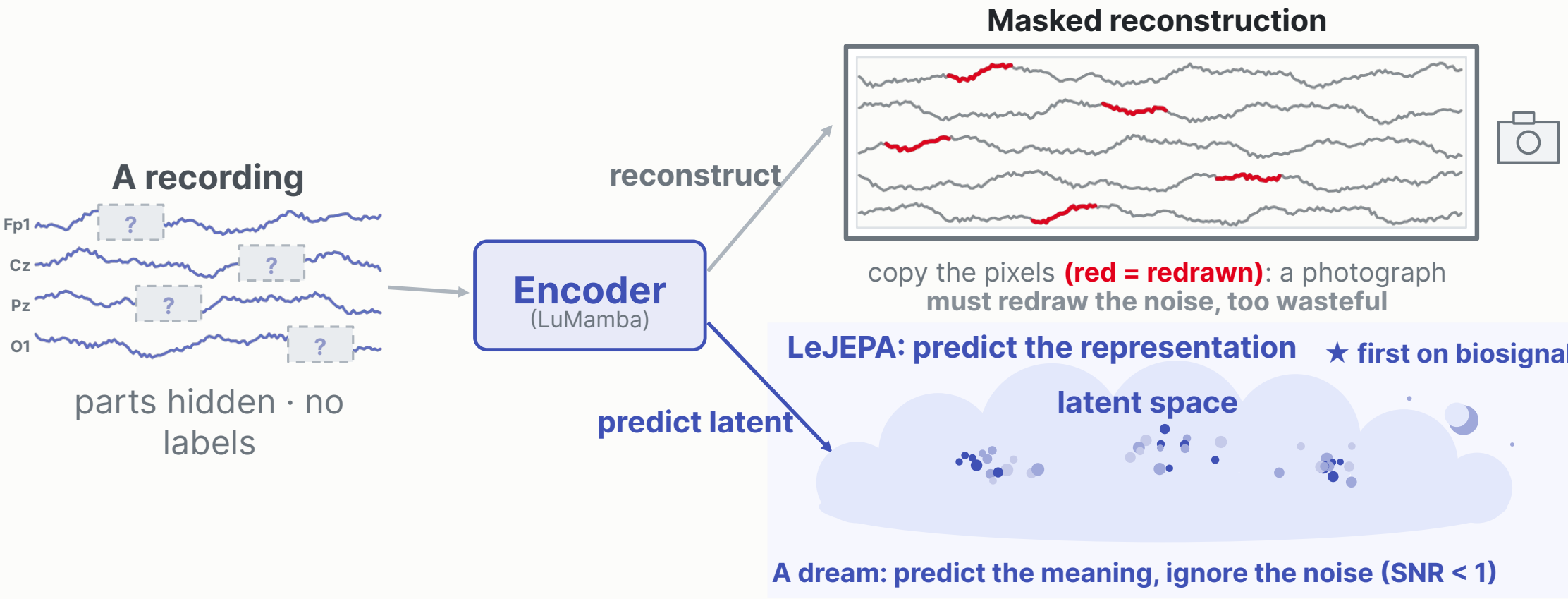




≈ 21,000 h of EEG
almost none labeled

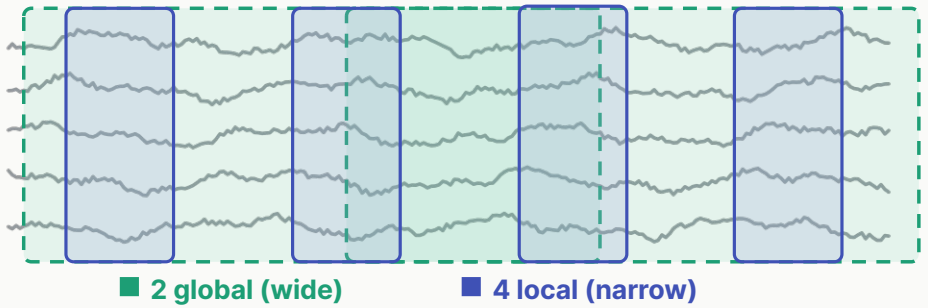
How do we train this? → *World...*

Train it by predicting the hidden part: the meaning, not the pixels.



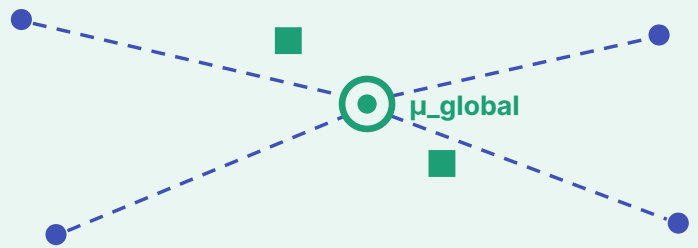
A short snippet should predict the whole window.

raw EEG window



LuMamba encoder

latent space, not raw

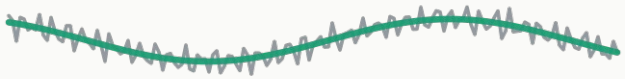


each local view predicts μ_global

$$\mathcal{L}\_JEPA = \frac{1}{N_{local}} \sum ||\mu_{global} - v_{local}||^2$$

underlying structure

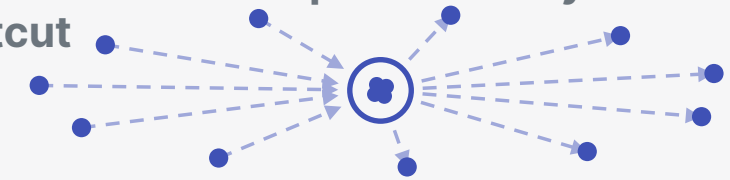
+ superficial noise



We learn the structure, not the noise.

↻ the signal is buried in noise (SNR < 1), only the structure is shared across views.

representation collapse: the easy shortcut

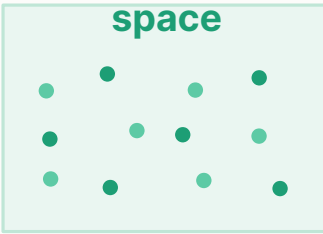


every embedding → one point · loss = 0, learns nothing

We need robust regularization

to keep it healthy:

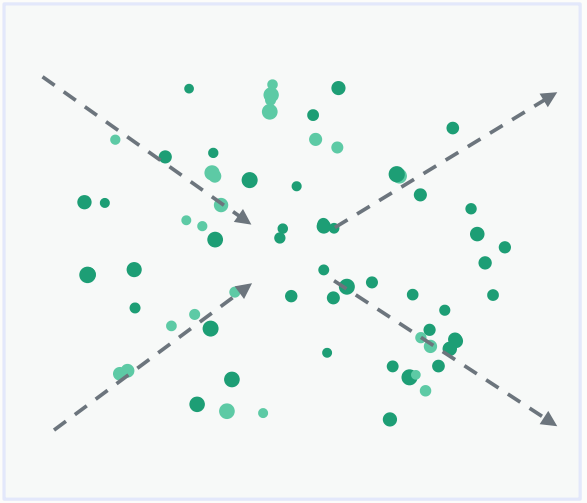
the embedding space



SIGReg: make every random shadow look Gaussian.

no teacher · no stop-gradient · no schedulers

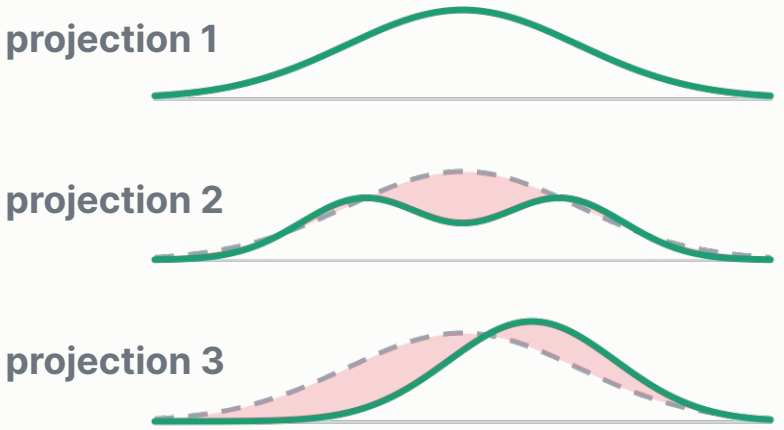
embedding cloud



project onto M = 300 random directions

project

each 1-D shadow vs a Gaussian



Epps–Pulley test — the **red deviation** from a Gaussian is penalized to zero

Why isotropic? Because classes stay separable in every direction

Every direction carries signal, a probe can split the classes whichever way they fall.

→ minimizes downstream risk · no collapse

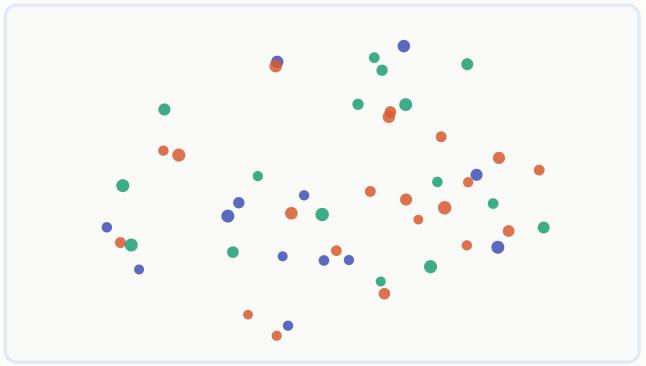
Reconstruction clusters, LeJEPA spreads, whereas mixed gets both.

t-SNE of learned embeddings · colored by class

Reconstruction-only

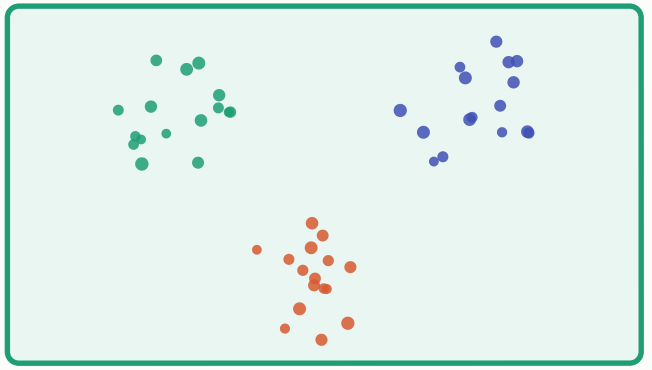
tidy clusters · brittle on new montages

LeJEPA-only



spread out · weak class structure

Mixed = reconstruction +



structured AND spread · best of both

Alzheimer's detection on an unseen 16-channel montage (APAVA): AUPR

0.77	0.82	0.97

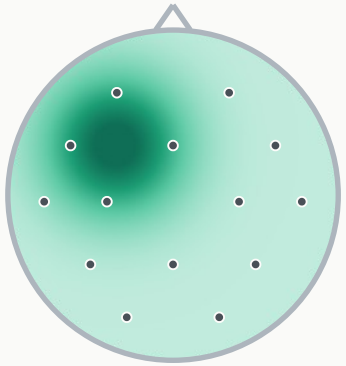
Mixed keeps clusters AND generalizes, +20% AUPR over reconstruction, state-of-the-art 0.97.

The objective also shapes where the model looks.

query attention on the scalp · one learned query

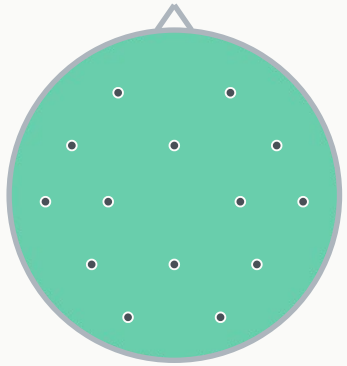
attention high

Reconstruction-



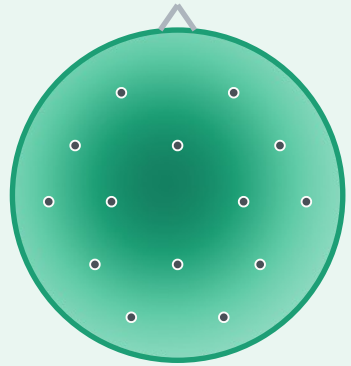
sharp · localized peak

LeJEPA-only



flat · homogeneous

Mixed



graded · spans regions

Mixed relaxes reconstruction's locality, **graded attention spans brain regions**, capturing higher-order patterns.

Across the clinical tasks · one model · 4.6M params:

abnormal EEG (TUAB)	artifacts (TUAR)	Parkinson's (TDBrain)	Alzheimer's (APAVA) SoA
81.0% bal. acc	**0.90 AUROC**	**0.96 AUROC**	**0.97 AUPR**

We taught a model to dream in EEG.

Tiny & topology-free

LUNA channel-unify + Mamba

4.6M any montage & length

World-model SSL for EEG

Predict the meaning (JEPA),

keep it healthy (SIGReg)

And it pays off

SoA Alzheimer's: 0.97 AUPR

best on unseen montages

Code & models — open source

github.com/pulp-bio/biofoundation

LUNA · FEMBA · PanLUNA · LuMamba

Thank you.

Thorir Mar Ingolfsson

thoriri@iis.ee.ethz.ch · ETH Zürich, Integrated Systems Laboratory