Contact us: akastrati@ethz.ch, lanzendoerfer@ethz.ch, rrigoni@ethz.ch, sjohn@ethz.ch, wattenhofer@ethz.ch

Paper:

Adapting Neural Audio Codecs to EEG

Ard Kastrati, Luca Lanzendörfer, Riccardo Rigoni, John Staib Matilla, Roger Wattenhofer
ETH Zurich, Distributed Computing Group

Motivation

- EEG foundation models need discrete representations for scalable learning
- EEG datasets are small, diverse in hardware, and harder to compress than audio
- Neural audio codecs already excel at high-fidelity tokenization — can we reuse them?

Main Idea

- Simply "**feed**" the EEG data to an Audio codec
- **Map** raw EEG into stride-based framing of DAC for direct encoder–decoder reuse
- **Fine-tuning** on EEG improves fidelity vs. training a codec from scratch
- **DAC-MC**: multi-channel extension with cross-channel attention + channel-specific decoding

Evaluation

- Tested on **TUAB Abnormal** & **TUEP Epilepsy** datasets
- Good **reconstruction fidelity** (spectrogram loss)
- **Downstream classification**: minor drop, but clinically relevant features retained

METHODS

AUDIO → EEG

Dealing with Sampling Rate

- Normalize EEG to audio scale (±200 µV → [-1, 1])
- Treat 512 samples = 1 token (≈1 s EEG vs. 13 ms audio)
- Direct inference with unmodified pretrained DAC gives coherent EEG outputs

Dealing with Multi Channels

- DAC-SC (single-channel): compress channels independently; simple, but ignores spatial correlations
- DAC-MC (multi-channel): joint encoding with cross-channel attention + channel-specific decoding
- Lightweight adapters → preserve audio-pretrained weights and reduce compute via channel grouping

Configuration Trade-offs

- **Sampling-rate**: higher rate → finer temporal detail vs. higher bitrate
- **Vocabulary size**: shrink codebooks to reduce bitrate (fine-tuning retains performance)
- **Residual depth**: adjust codebook layers for compression–fidelity balance (pruned vs. retrained variants)

DATASET

- TUH EEG corpus
  - fine-tune on full EEG,
  - evaluate on TUAB (abnormal) & TUEP (epilepsy)

ARCHITECTURE

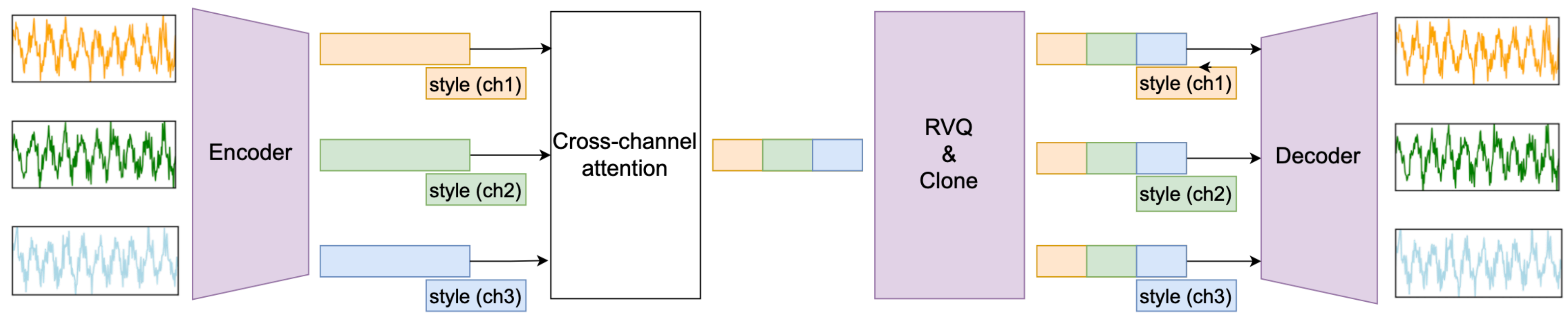

Figure 1: DAC-MC. Purple modules form the pretrained DAC backbone.

QUALITATIVE RESULTS

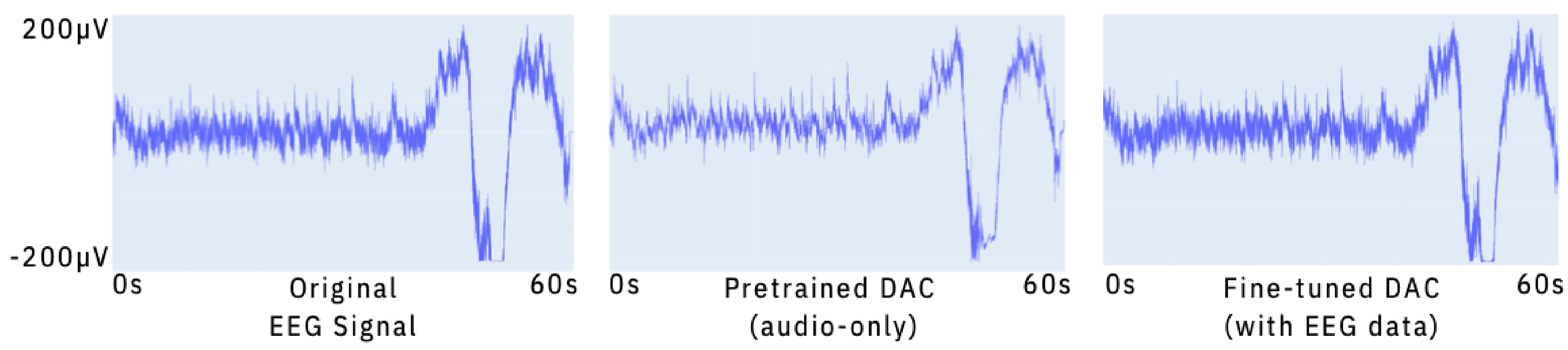

Figure 2: Example reconstruction with audio-pretrained codec and fine-tuned codec with EEG data.

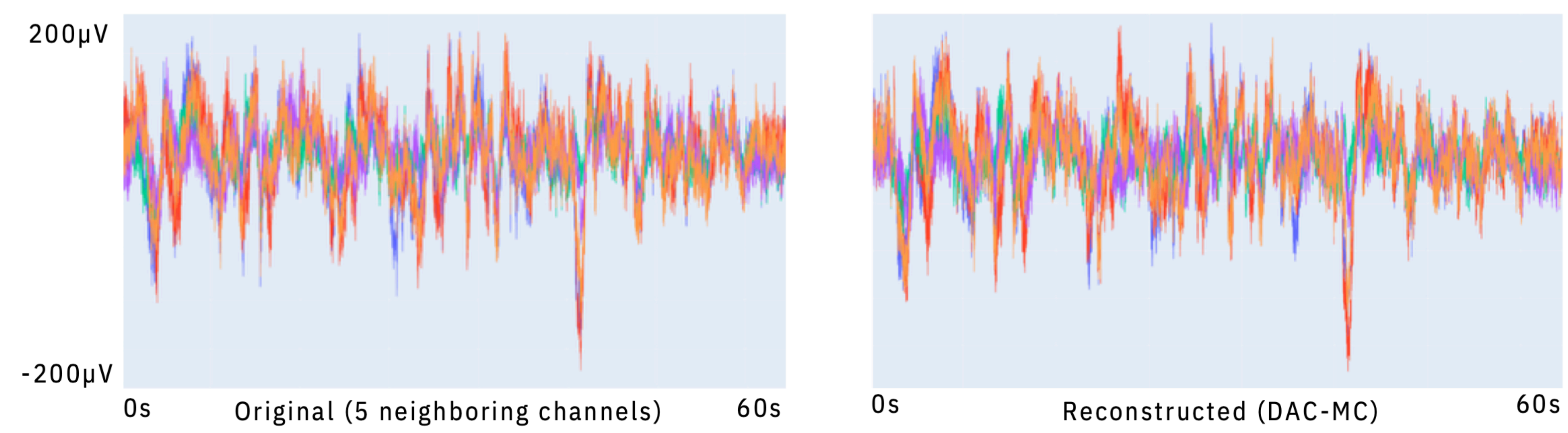

Figure 3: Example reconstruction with fine-tuned codec with multi-channels.

RESULTS

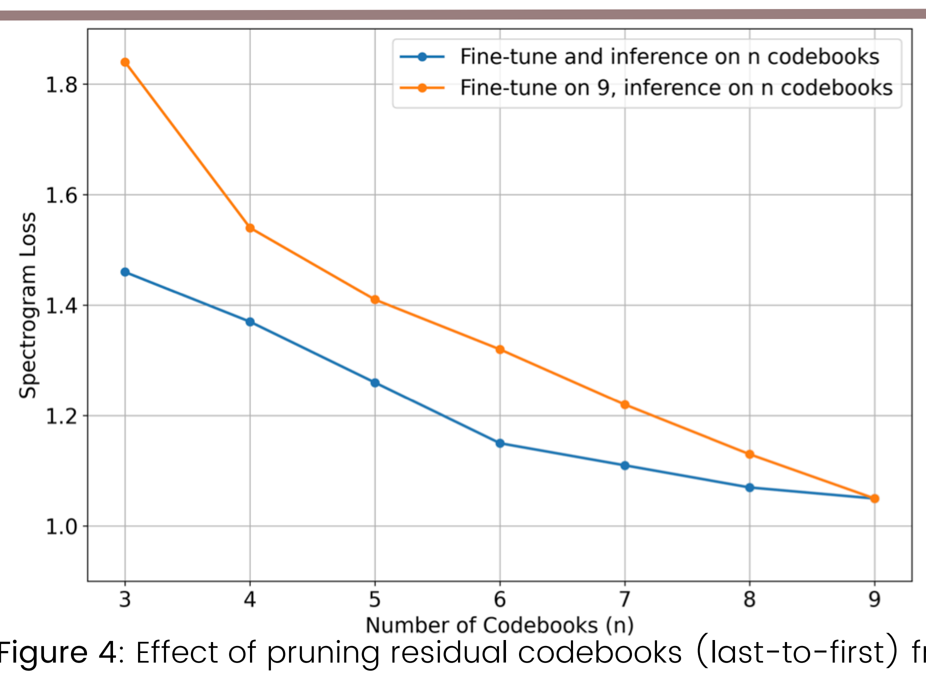

Figure 4: Effect of pruning residual codebooks (last-to-first) from the vector quantizer before and after fine-tuning the Audio-to-EEG model. After: the model is fine- tuned with all nine codebooks and, at inference, only the first n are kept. Before: the model is fine-tuned from scratch using only the first n codebooks.

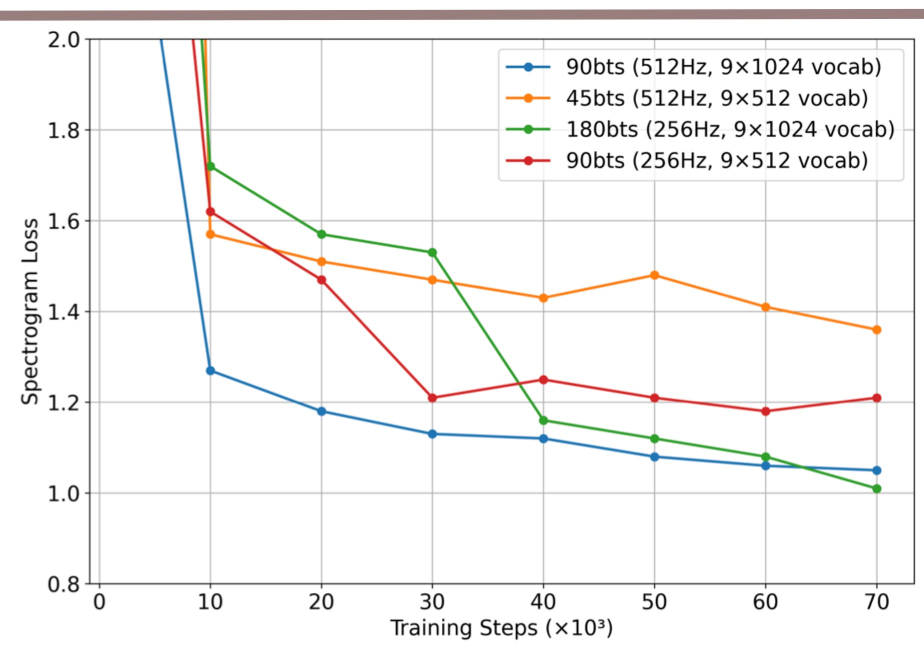

Figure 5: Effect of upsampling and alphabet size reduction on reconstruction fidelity.

Model	Loss
Audio-Pretrained	2.50
Scratch	1.46
Audio-to-EEG	1.05

Table 1: Comparison of the spectrogram reconstruction losses from test set showing that fine-tuning an audio-pretrained codec for EEG yields the best performance.

Mode	Epilepsy	Abnormal
Single-Channel (DAC-SC)	80 %	83 %
Single-Channel (DAC-MC)	82 %	81 %
Random-Groups (DAC-MC)	85 %	78 %
Manual-Groups (DAC-MC)	85 %	78 %
Baseline	84 %	82 %

Table 2: Benchmark accuracy for Epilepsy and Abnormal EEG datasets. DAC-SC is a single-channel model. DAC-MC is a multi-channel model that can be used either per-channel decoding ("Single-Channel (DAC- MC)") or jointly on multiple channels with groups chosen at random or manually ("Random-Groups"/"Manual-Groups" (DAC-MC)). The Baseline is the best accuracy obtained when training and testing on the original signals.

CONCLUSIONS

- **Pretrained vs Scratch**: Fine-tuning DAC on EEG lowers spectrogram loss (≈1.05) vs. Scratch (1.46) and Audio-Pretrained (2.5)
- **Codebook trade-offs**: Reducing residual codebooks (9 → 6) slight increases loss (<10%), extreme pruning (9 → 3) doubles it
- **Upsampling** (256 → 512 Hz): minor gains in temporal detail, but much higher compute cost
- **Alphabet reduction** (1024 → 512 entries): degrades reconstruction unless combined with pruning + fine-tuning
- **Multi-channel vs. single-channel**: DAC-MC with grouped channels boosts epilepsy detection (85% vs. 80%) but per-channel decoding preserves abnormal EEG features better (81% vs. 78%)