Quantizing Quakes: Exploring Neural Codecs for Earthquake Signal Analysis

Archisman Bhattacharjee[a], Pawan Bharadwaj[a]

[a] Centre for Earth Sciences - Indian Institute of Science, Bangalore, India

Introduction

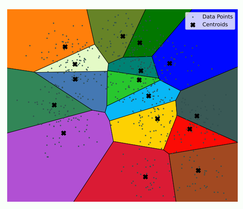
Neural codecs have demonstrated significant success in audio processing by transforming continuous waveforms into discrete tokens. Originally developed to reduce transmission latency, these tokenized representations have proven particularly valuable in enabling natural language processing-inspired approaches to continuous signal data, while also providing efficient storage solutions for large datasets. Recent models like Meta's AudioGen[2] and MusicGen[3] have shown how such tokenized representations can serve as effective building blocks for complex audio generation and manipulation tasks. We introduce EQcodec, an adaptation of Meta AI's Encodec[1] architecture for seismic signal analysis. Using residual vector quantization, our model converts seismic waveforms into discrete tokens while preserving essential signal characteristics across different quantization levels. Our results demonstrate the model's ability to encode seismic signals at varying levels of discretization, each representing different degrees of information content. Initial experiments indicate these compressed representations can improve performance on analytical tasks while significantly reducing storage requirements. These preliminary results motivate further investigation into new analytical approaches in seismological data analysis, such as transformer-based architectures for time series analysis and predictive deconvolution.

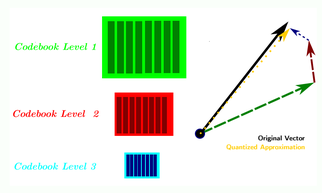
Residual Vector Quantization

Vector quantization (VQ) is a technique that maps continuous-valued vectors to a finite set of representative values. Residual Vector Quantization (RVQ) extends this concept by quantizing data in multiple stages using codebooks - collections of reference vectors called centroids. The process works as follows:

- In the first stage, input vector x is approximated by its nearest centroid: $\hat{x}_1 = q_1(x)$
- The first residual $r_1 = x - \hat{x}_1$ is quantized using a second codebook: $\hat{x}_2 = q_2(r_1)$
- Subsequent stages follow the pattern $r_i = r_{i-1} - \hat{x}_i$, each refining the previous approximation
- The final reconstruction is the sum of all quantized vectors: $\hat{x} = \sum_i \hat{x}_i$

This sequential refinement strategy allows RVQ to achieve high compression rates while maintaining good reconstruction quality, making it particularly effective for neural compression tasks.

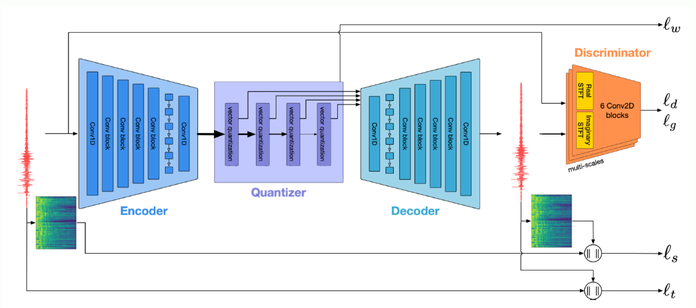




The EQcodec Architecture

The EQcodec architecture is a direct adaptation of Meta AI's Encodec model, originally designed for audio compression, now repurposed for seismic waveform compression. The architecture consists of an encoder-decoder framework with a quantizer in between. Key components and parameters:

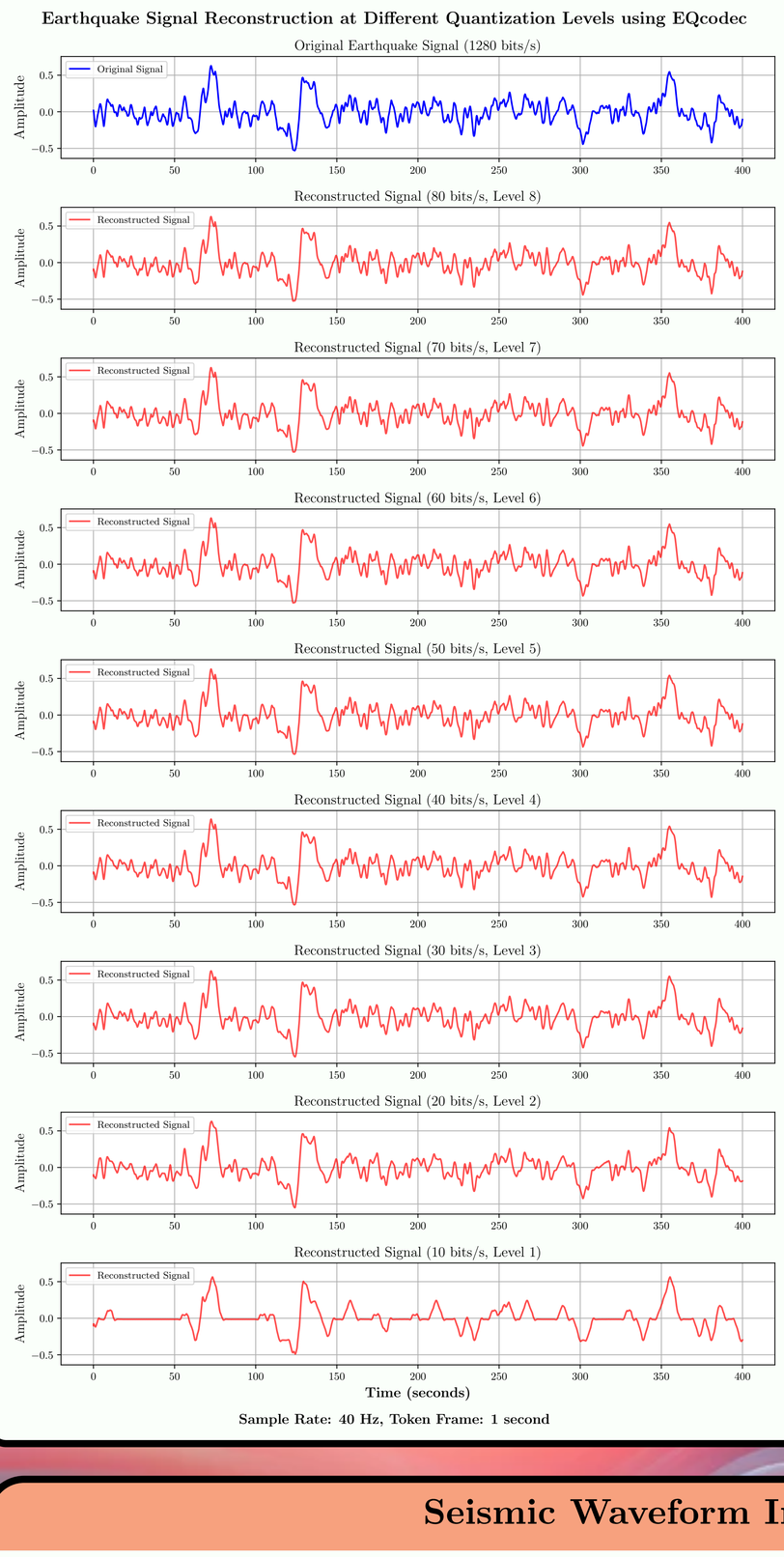
- Encoder: Uses strided convolutions followed by LSTM layers
- Decoder: Uses LSTM layers followed by transposed convolutions
- Downsampling: Total downsampling factor of 40 between input and latent space
- Dimension: Uses 32-dimensional latent space for encoding seismic features
- Quantization: Implements Residual Vector Quantization (RVQ) with 8 codebooks and 1024 centroids per codebook
- Sampling Rate: Trained over 40Hz seismic data



The discriminator uses a multi-scale STFT approach with three parallel discriminators operating at FFT sizes of 128, 64, and 32 points to ensure robust adversarial training across various frequency scales of the seismic signal. The model is trained using multiple objectives:

- Reconstruction losses (ℓ_t and ℓ_s) in both time and spectral domains
- Adversarial losses for generator (ℓ_g) and discriminator (ℓ_d)
- Commitment loss (ℓ_w) from the residual vector quantization

Earthquake Data Compression with EQcodec: Demo



- Training and testing were performed on single-channel earthquake signals sampled at 40 Hz, with all component traces from 49.8 GiB of training data and 13.1 GiB of test data split into individual channels.
- Figure on left demonstrates how EQcodec progressively compresses seismic waveforms through residual vector quantization using 8 codebooks on a randomly chosen test set trace.
- The original signal (top) is sampled at 40 Hz with 32-bit precision, yielding 1280 bits/s.
- Below are reconstructions using progressively fewer codebooks, with bit rates decreasing from 80 to 10 bits/s (levels 8 to 1, corresponding to the number of codebooks used).
- For every 1-second segment (40 samples), there is a token frame, the number of tokens being the number of codebook levels.
- The visualization reveals how different quantization levels balance data reduction against signal fidelity in the encoded seismic waveforms.

Reconstruction Error Metrics across Quantization Levels

Compression Level	MSE (Mean ± Std) $\times 10^{-3}$	MAE (Mean ± Std) $\times 10^{-2}$
1	16.1 ± 125.6	82.3 ± 52.6
2	4.56 ± 97.5	35.2 ± 37.1
3	2.75 ± 87.1	22.5 ± 32.4
4	2.11 ± 79.1	17.2 ± 29.8
5	1.83 ± 75.3	14.1 ± 28.7
6	1.66 ± 69.8	12.3 ± 27.6
7	1.57 ± 66.9	11.2 ± 27.1
8	1.51 ± 64.8	10.5 ± 26.9

Compression Ratio as a Function of Quantization Level

Codebook Levels	Bits per Frame	Bits per Second	Compression Ratio
1	10	10	128.00
2	20	20	64.00
3	30	30	42.67
4	40	40	32.00
5	50	50	25.60
6	60	60	21.33
7	70	70	18.29
8	80	80	16.00

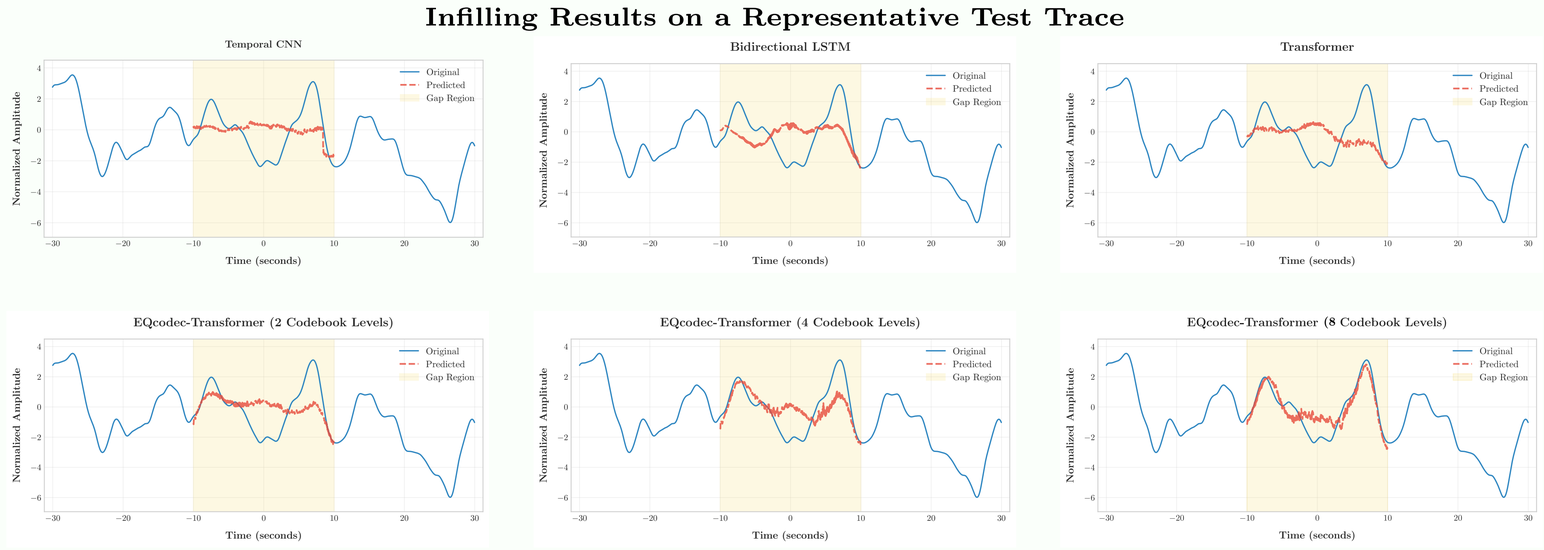
Seismic Waveform Infilling Experiments

We design a waveform infilling task, which requires understanding signal context to reconstruct missing portions, to evaluate learned seismic representations. We compare our quantized approach against traditional baselines including Temporal CNN, Bi-LSTM, and vanilla Transformer architectures. Unlike the baselines that work directly with raw waveforms, the EQcodec variants first encode the seismic data into discrete embeddings using different quantization levels (2, 4, and 8 codebooks) before feeding them to the Transformer model. Results show progressive improvement in reconstruction quality with increasing number of codebooks, with the 8-level EQcodec achieving the best performance (MSE: 0.616, MAE: 0.484). Visual comparisons demonstrate superior reconstruction of the masked regions, particularly in preserving the waveform characteristics.

Test Set Performance Metrics of Different Models in Infilling Task

Model	MSE Loss	MAE Loss
Temporal CNN	1.560	0.836
Bi-LSTM	0.983	0.663
Transformer (TF)	0.987	0.659
EQcodec (2 Levels) + TF	0.854	0.604
EQcodec (4 Levels) + TF	0.748	0.548
EQcodec (8 Levels) + TF	0.616	0.484

Infilling Results on a Representative Test Trace



Practical Implications and What's Next?

- Modern seismic networks generate massive datasets that strain storage and network infrastructure
- Our approach enables efficient storage and transfer while preserving signal quality
- Makes large-scale seismic analysis accessible to institutions with limited resources
- Compressed representations can be effectively integrated into deep learning workflows
- Create an open-source EQcodec trained on large-scale global seismic datasets, enabling widespread adoption within the seismological community
- Explore applications of recent deep learning advances to seismological analysis through these tokenized representations, such as predictive deconvolution and other analysis tasks.

References

1. Défossez, A., Copet, J., Synnaeve, G., & Adi, Y. (2022). High fidelity neural audio compression. arXiv preprint arXiv:2210.13438.

2. Kreuk, F., Synnaeve, G., Polyak, A., Singer, U., Défossez, A., Copet, J., ... & Adi, Y. (2022). Audiogen: Textually guided audio generation. arXiv preprint arXiv:2209.15352.

3. Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., ... & Défossez, A. (2024). Simple and controllable music generation. Advances in Neural Information Processing Systems, 36.