Towards a Foundation Model for Fusion: Multimodal Representation Learning of Plasma State and Control

Kouroche Bouchiat[1] *Nathaniel Chen*[1] *Peter Steiner*[1] *Azarakhsh Jalalvand*[1] *Egemen Kolemen*[1,2]

[1]Princeton University [2]Princeton Plasma Physics Lab

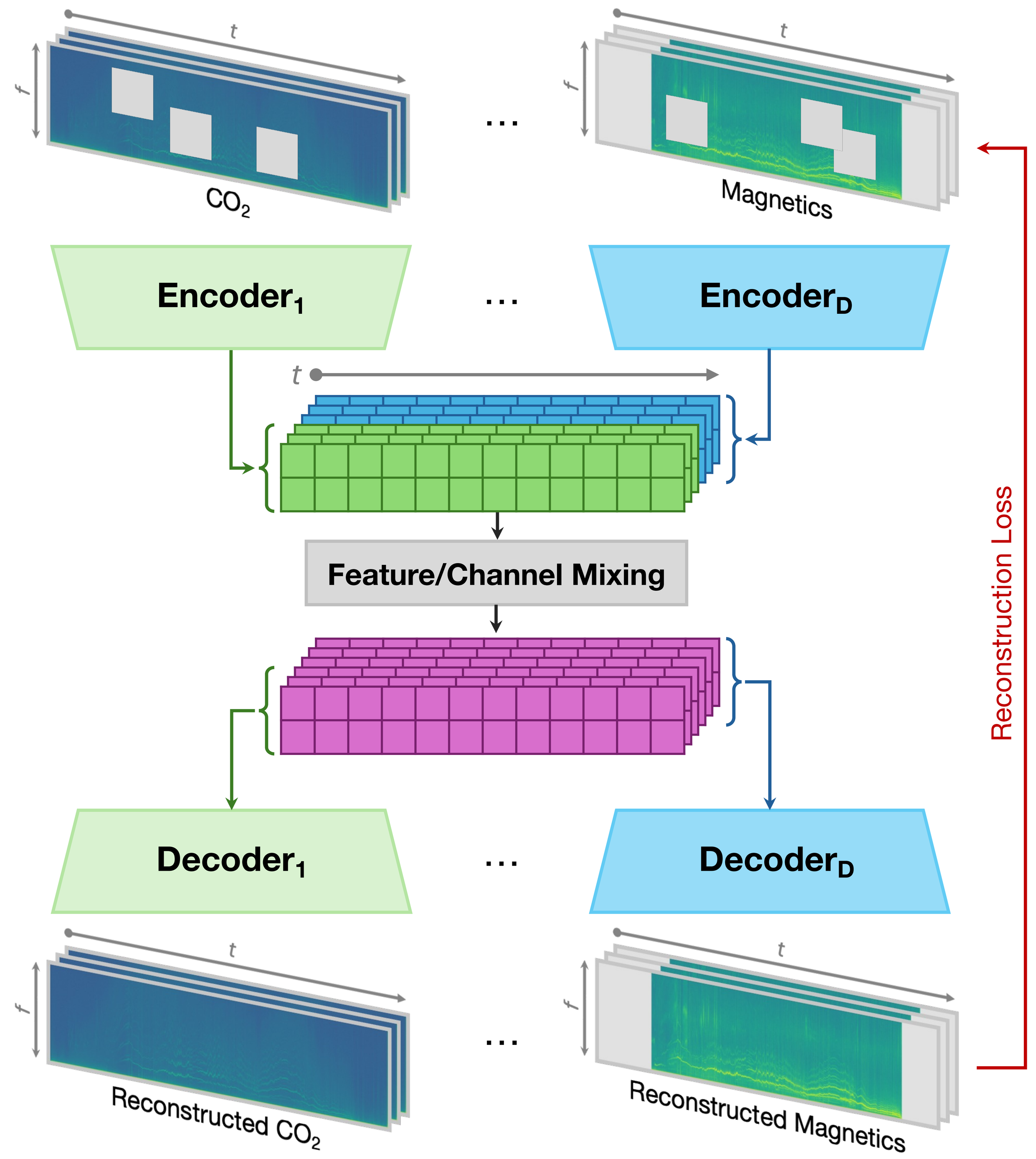
Motivation

- **Foundation models:** Large neural nets pre-trained on diverse data to learn general-purpose representations adaptable to multiple downstream tasks.
- **Diagnostic fusion:** Building on the success of *Diag2Diag* [1], we seek to integrate multiple diagnostics to build a foundation model for fusion.
- **Self-supervised learning** is able to leverage large-scale multi-modal data to learn a universal representation of plasma dynamics.

Dataset

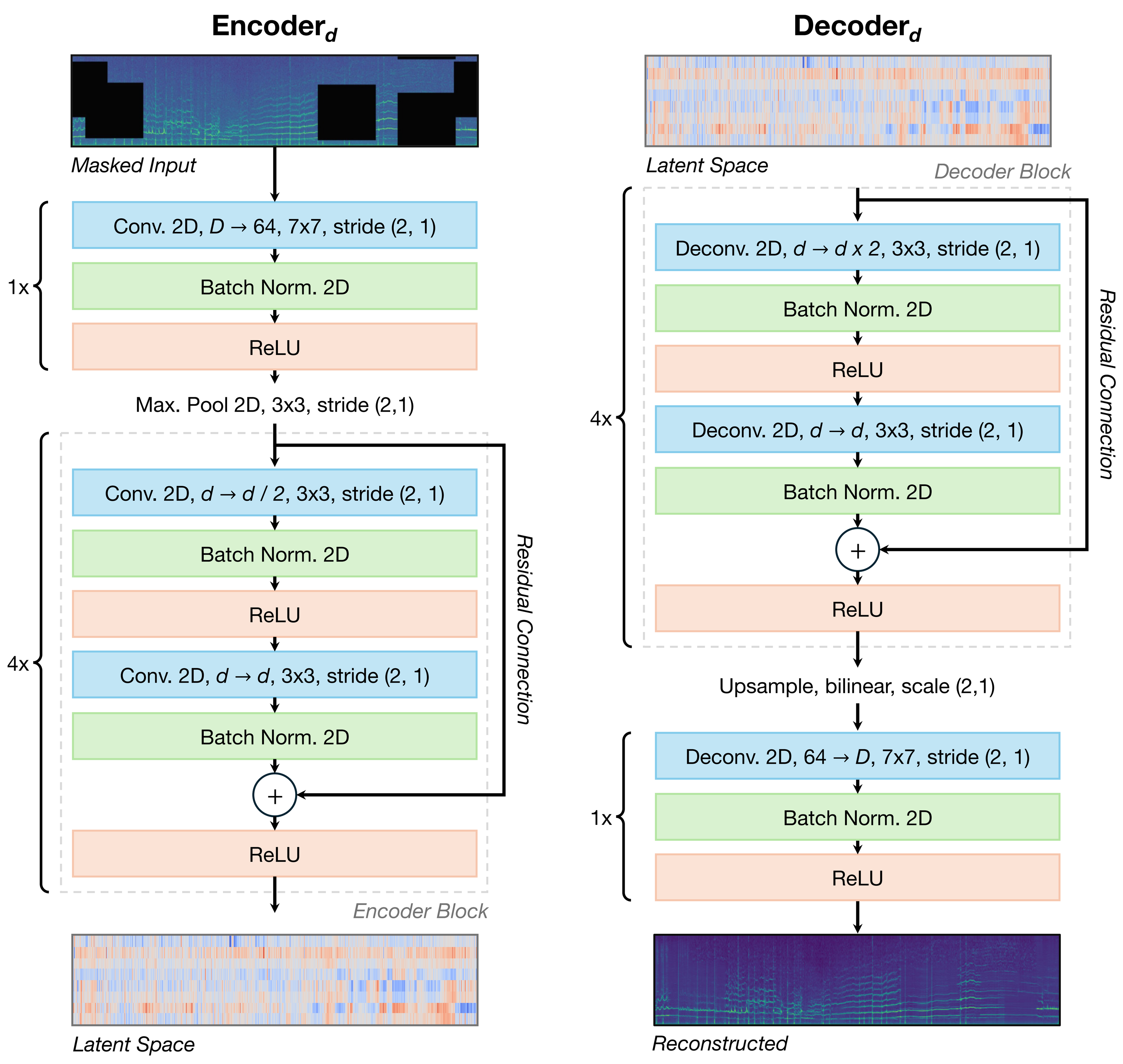
- Currently **~2500 DIII-D shots** used to train with 3 high-res. diagnostics:
  - **Electron Cyclotron Emission:** 40 channels @ 500 kHz, edge-to-core.
  - **CO2 Interferometry:** 4 chords @ 1.6 MHz, 3 vertical and 1 radial.
  - **High-Speed Magnetic Probes:** 8 channels @ 2 MHz.
- **Resampled** to 500 kHz and **re-aligned** in time.
  - Converted to spectrograms using shared STFT parameters.

Architecture



- Each diagnostic is assigned an **encoder/decoder pair**.
  - Encode/reconstruct the spectrogram to/from its **latent representation**.
- During training ~75% of spectrogram is **masked out** and **reconstructed**.

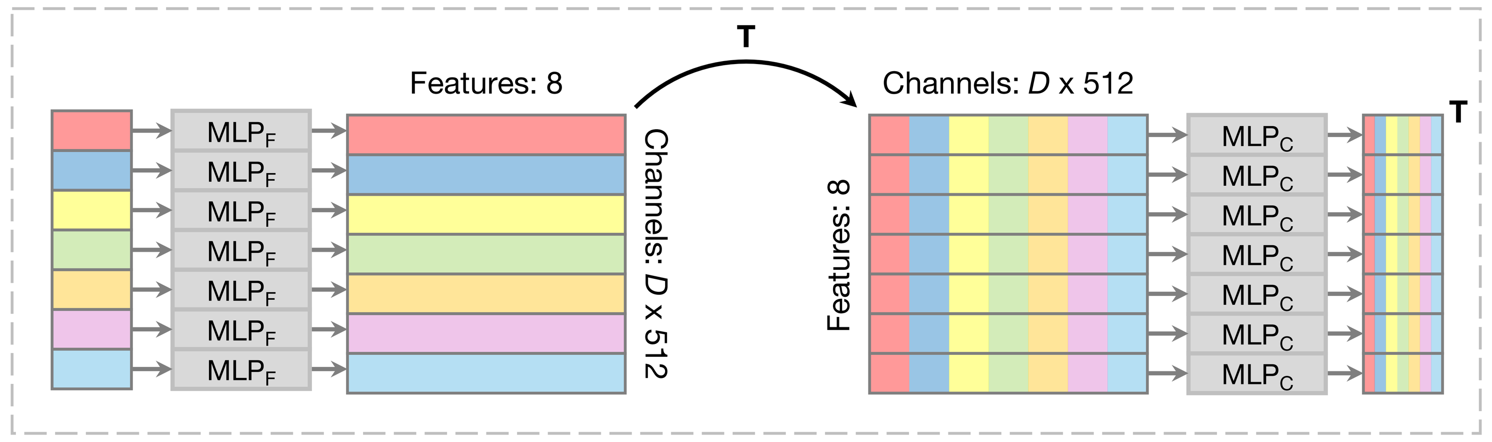
Encoding & Decoding



- Down-/Upsample in frequency dimension using **2D convolution** blocks.
  - Encoder based on *ResNet* [2] backbone, decoder mirrors encoder.
- Trained using Mean Squared Error (MSE) on the unmasked spectrograms.

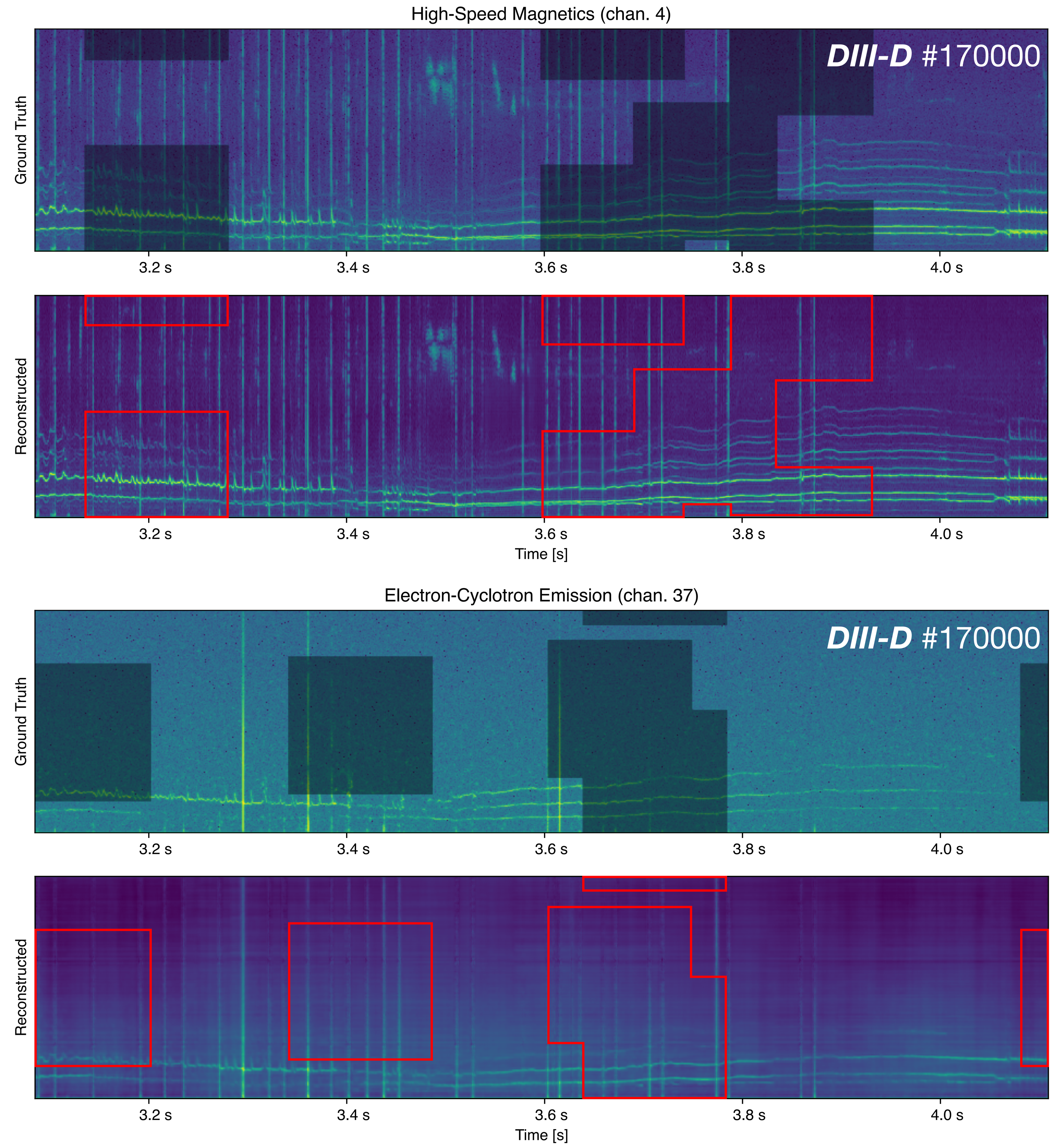
Latent Space Mixing

Feature/Channel Mixing



- Inspired by *MLP-Mixer* [3], replace convolution/attention with simple MLPs that **mix information** along different dimensions separately.
- First mix across **frequency bins**, then mix across **diagnostic channels**.
  - Single linear layer per dimension, avoiding quadratic parameter growth.
- **Channel mixing** enables information sharing across different diagnostics, learning how to combine their latent representations.

Example Reconstruction



- **Shaded regions** above are masked out before being input.
- **Red regions** below are the same regions reconstructed by the model.

Next Steps

- Scale dataset to data available at DIII-D (~20h of ~300h available so far.)
  - Add support for larger number of diagnostics/actuators.
  - Add support for mixed sampling rates.
- Investigate and adapt latent space for downstream tasks.
  - Mode identification, profile reconstruction, tearing mode prediction, etc…

Conclusion

We demonstrated that self-supervised learning on fusion diagnostic spectrograms enables robust reconstruction of masked regions, establishing a pathway toward a general-purpose foundation model for fusion.

This work was made possible by the Department of Energy under *DE-FC02-04ER54698*, *DE-SC0024527* and Princeton Laboratory for Artificial Intelligence under *Award 2025-97*.

[1] Jalalvand, A. et al. "*Multimodal super-resolution: discovering hidden physics and its application to fusion plasmas,*" Nat Commun. 16, 8506 (2025) [2] K. He et al. "*Deep Residual Learning for Image Recognition,*" CVPR 2016 [3] Tolstikhin, I. et al. "*MLP-Mixer: An all-MLP Architecture for Vision,*" NeurIPS 2021