SUNet: A Spectral UNet for Axion Dark Matter Signal Denoising

Akbota Assan[1] **Aobo Li**[1]

[1]Rare AI Lab, Halıcıoğlu Data Science Institute, UC San Diego

aassan@ucsd.edu aol002@ucsd.edu

Abstract

Axion dark matter searches such as ABRACADABRA produce continuous high-rate time series in which injected signals occupy a single narrow frequency bin per time frame, a structure that is natural in the spectral domain but opaque in the raw time domain. Existing TIDMAD denoising approaches either operate directly on raw time series or require separate model weights per frequency band, limiting generalizability and deployment practicality. We introduce SUNet, a Spectral UNet that operates on real and imaginary Short-Time Fourier Transform (STFT) components, enabling inverse-STFT signal reconstruction while training a single weight set across the full axion frequency range. SUNet is the second model on TIDMAD, after WaveNet, to generalize across the full frequency range without per-band retraining, and the first to do so in the spectral domain. Through frequency-resolved evaluation, we show that SUNet achieves strong recovery from 20 kHz to 4 MHz while degrading below 20 kHz, consistent with elevated 1/f noise and STFT resolution limits near DC. This characterization, absent from prior TIDMAD evaluations, reveals frequency-dependent behavior that aggregate benchmark scores obscure and that matters for targeted axion mass searches.

1 Introduction

The axion is a well-motivated dark matter candidate whose detection requires identifying an extremely weak oscillating signal buried in broadband electromagnetic noise. Experiments such as ABRACADABRA at MIT [2] operate by continuously recording SQUID magnetometer data at high sampling rates ($f_s = 10$ MHz), injecting a known reference signal via a signal generator, and attempting to recover that signal after denoising. The TIDMAD benchmark [1] formalizes this problem: given a noisy time series (channel 1, SQUID) and a reference signal (channel 2, signal generator), produce a denoised output whose signal-to-noise ratio at the injected frequency is maximized, scored as

$$S \;=\; \log_{5.27}\left(\frac{\mathbf{SNR}_{\mathrm{sg,norm}} \cdot \mathbf{SNR}_{\mathrm{squid}}}{N}\right), \tag{1}$$

where $\mathbf{SNR}_{\mathrm{sg,norm}}$ is the per-second reference SNR vector normalized by its global maximum, $\mathbf{SNR}_{\mathrm{squid}}$ is the per-second denoised-channel SNR at the injected frequency, and N is the total number of seconds evaluated.

A key physical property of axion signals constrains the problem structure: for a given axion mass m_a, the signal oscillates at a single frequency $f_a = m_a c^2/h$, occupying exactly one narrow frequency bin per time frame in the STFT representation. In the time domain, this tone appears as a global oscillation indistinguishable from noise without spectral analysis. The STFT makes the structure explicit: the signal is spatially localized in the time-frequency plane, and denoising reduces to recovering one bright bin per frame from a noisy spectrogram.

Despite this natural spectral structure, all existing TIDMAD deep learning baselines (FCNet, PUNet, Transformer, RNN) operate on raw time-domain segments, producing 256-class per-timestep classification outputs trained with Focal Loss [5]. The TIDMAD authors note that all architectures except WaveNet fail to generalize across the full frequency range with a single weight set, requiring four separately trained specialist models per architecture; at inference, each validation file is denoised by the specialist trained on its corresponding frequency range [1]. WaveNet [4] is the sole exception, generalizing across the full range as a single time-domain generalist model. All published TIDMAD evaluations report a single aggregate benchmark score, collapsing performance across two orders of magnitude in frequency into one number and leaving unanswered which specific axion mass ranges each model serves well.

We make the following contributions:

- **SUNet**, a spectral-domain UNet denoiser for TIDMAD that exploits the narrow-band structure of axion signals in the time-frequency plane.
- A **two-channel real/imaginary STFT formulation** that retains phase information and enables direct inverse-STFT waveform reconstruction without a separate phase-estimation step.
- A **single-weight generalist model** trained across the full axion frequency range, avoiding the per-band retraining and frequency-routing required by specialist baselines.
- A **frequency-resolved evaluation** across all 20 TIDMAD validation files, showing that aggregate benchmark scores obscure substantial frequency-dependent variation in denoising performance.

2 TIDMAD Benchmark and Dataset

The TIDMAD dataset [1] consists of 20 training files and 20 validation files, each containing 200 seconds of dual-channel time series recorded at 10 MHz sampling. Each file stores approximately 2×10^9 samples per channel as 8-bit integers, yielding roughly 2.2 GB per file and an 88 GB dataset footprint. The injected signal frequency f_a steps through 309 values from 1.1 kHz to 4.9 MHz, with each validation file covering a contiguous sub-range of approximately 15 frequencies. Within each file, the injected frequency advances every approximately 12 seconds.

The benchmark fine score (Eq. 1) aggregates per-second SNR measurements across all files into a single scalar. For each second, the peak of channel 2's power spectral density (PSD) identifies the injected frequency f_a via a sharpness-based peak finder (`findPeak`); the SNR of channel 1 (the denoised SQUID output) is then measured at that same frequency bin. The global normalization $\mathbf{SNR}_{\mathrm{sg,norm}} = \mathbf{SNR}_{\mathrm{sg}} / \max(\mathbf{SNR}_{\mathrm{sg}})$ accounts for the detector's frequency-dependent hardware response: seconds where the hardware bandpass filter naturally attenuates the reference signal receive near-zero weight in the aggregate score, regardless of denoising quality. This means the aggregate score can appear reasonable even when a model fails at specific frequency ranges that are down-weighted by the hardware response.

The published FCNet baseline achieves a fine score of 6.43 via an ensemble of four per-band specialist models. WaveNet achieves the highest published score as a single-model generalist [1]. FCNet, PUNet, Transformer, and RNN all operate in a per-band split regime by default. WaveNet is the only prior architecture that generalizes across the full frequency range with a single weight set [1].

3 SUNet

SUNet processes 1-second segments ($L = 10{,}000{,}000$ samples at 10 MHz) via an STFT frontend with $N_{\mathrm{FFT}} = 40{,}000$ and hop $H = 20{,}000$, giving frequency resolution $\Delta f = 250$ Hz/bin and $N_F = 20{,}001$ frequency bins. The real and imaginary parts of the complex spectrogram are stacked as two input channels, yielding $\mathbf{X} \in \mathbb{R}^{2 \times N_F \times T}$ with $T = 501$ time frames. This two-channel decomposition preserves full complex phase information: discarding the imaginary channel, as in magnitude-only spectral approaches, would require phase estimation at reconstruction time and introduce approximation error. By predicting both channels jointly, the network preserves the phase relationships required for iSTFT reconstruction.

The axion signal at frequency f_a concentrates energy in a single bin $k_a = \lfloor f_a/\Delta f \rceil$, making denoising a geometrically localized operation in the spectral domain. The network is a 2D UNet [3] operating jointly on frequency and time axes, with a channel ladder $C = [16, 32, 64, 128, 256]$, Conv2d-GroupNorm [6]-GELU blocks, and an asymmetric pooling strategy that applies $\mathrm{MaxPool}_{1\times 2}$ (time-only) at the first encoder stage to preserve all 20,001 frequency bins before any spatial compression, followed by $\mathrm{MaxPool}_{2\times 2}$ at stages 2 to 4. The output head predicts $\hat{\mathbf{X}} \in \mathbb{R}^{2\times N_F \times T}$, which is recombined into a complex spectrogram and inverted via iSTFT to recover the denoised time series. Full architecture equations and training details are given in Appendix A.

4 Experiments and Results

4.1 Aggregate Benchmark Score

Table 1: Aggregate fine denoising scores on the TIDMAD validation set. SUNet and WaveNet operate under the stricter single-generalist constraint; FCNet uses four per-band specialist models.

Model	**Regime**	**Fine Score** S
FCNet [1]	4 specialists, 4 bands	6.43
WaveNet [1]	Single generalist, time-domain	4.99
SUNet (ours)	**Single generalist, spectral**	**2.98**

SUNet trails the ensemble FCNet baseline on the aggregate score. This is expected: specialist models have a frequency-specific advantage at every frequency, with each sub-model optimized for its slice of the range. SUNet operates under a stricter constraint (one weight set, no frequency-specific optimization) and is more fairly compared against WaveNet as the only other published single-generalist model. The aggregate score is further limited as a comparison because its global normalization down-weights low-frequency performance, artificially narrowing the apparent gap.

4.2 Frequency-Resolved Signal Recovery

To reveal structure hidden by the aggregate score, we introduce a per-file frequency-resolved evaluation. We score each of the 20 validation files independently: the reference SNR vector is normalized by its maximum within that file only, rather than the global maximum across all files. This treats every frequency sub-range equally regardless of hardware sensitivity, producing one score per file that reflects average weighted signal recovery across that file's injected frequency sub-range. We apply the same evaluation to the official WaveNet single-weight checkpoint [4] obtained from the TIDMAD authors, enabling a direct frequency-resolved comparison between the two single-generalist models (Figure 1).

Table 2: Per-file score summary by frequency regime for both single-generalist models.

Regime	**Files**	**WaveNet**	**SUNet**
Below 20 kHz	0000–0005	-13.85 to $+2.48$	-2.77 to -1.18
20 kHz to 4 MHz	0006–0019	-2.35 to $+6.49$	$+2.33$ to $+4.23$
Mean (files 06–19)		4.12 ($\sigma = 2.82$)	3.52 ($\sigma = 0.54$)

Low-frequency degradation. Both models fail below 20 kHz, but with very different severity. SUNet scores range from -2.77 to -1.18, reflecting gradual degradation attributable to STFT resolution limits near DC and elevated $1/f$ noise: at $f_a = 5$ kHz the signal occupies bin $k_a = 20$, only 20 bins from DC where offset artifacts contaminate the lowest spectrogram entries, and negative scores extending to 9.5 kHz ($k_a = 38$) confirm that the $1/f$ noise floor remains elevated throughout the sub-20 kHz regime. WaveNet's low-frequency failures are far more severe: scores of -13.85 at 6–8 kHz and 9.5 kHz indicate degradation in signal recovery, suggesting the time-domain model cannot generalize its learned denoising transform to the highly structured low-frequency noise regime.

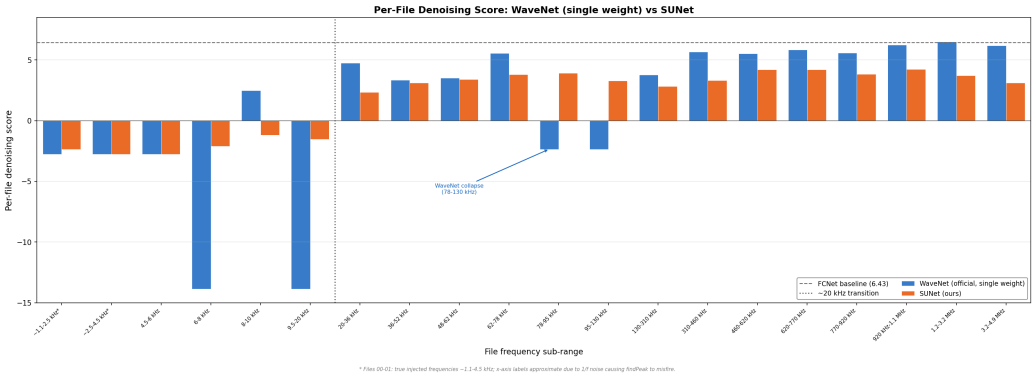

Figure 1: Per-file denoising score for WaveNet (blue) and SUNet (orange) across all 20 validation files using within-file normalization. Both use a single weight set. Scores above zero indicate positive signal recovery. Files 00–01 (asterisked): true injected frequencies ≈1.1–4.5 kHz; x-axis labels unreliable due to $1/f$ noise causing `findPeak` to misfire.

Mid-to-high frequency comparison. Above 20 kHz, WaveNet achieves higher peak scores (up to 6.49 vs SUNet's 4.23), reflecting the advantage of time-domain modeling at high frequencies where the noise floor is flat and the signal is strong. However, WaveNet exhibits an unexpected collapse at 78–130 kHz (files 10–11, scores -2.35), a sharp performance discontinuity that SUNet does not show. SUNet's mean score of 3.52 ($\sigma = 0.54$) across files 06–19 is lower than WaveNet's 4.12 but substantially more consistent ($\sigma = 2.82$ for WaveNet). For a real axion search targeting an unknown mass, consistency across the frequency range may be more operationally valuable than higher peak performance at select frequencies.

5 Discussion and Conclusion

We introduced SUNet, a spectral-domain denoiser on the TIDMAD axion dark matter benchmark. The central design choice is a two-channel real/imaginary STFT representation that preserves full phase information and enables iSTFT reconstruction, avoiding the phase estimation artifacts that afflict magnitude-only spectral approaches. This representation makes the single-bin-per-frame axion signal structure geometrically explicit, enabling a 2D UNet to generalize across the full frequency range without per-band retraining.

Frequency-resolved evaluation, itself a novel contribution to TIDMAD methodology, reveals a nuanced picture when comparing SUNet and WaveNet as the two single-generalist models. WaveNet achieves higher peak scores at high frequencies but exhibits negative denoising scores at specific frequency ranges: scores of -13.85 below 10 kHz and an unexpected collapse at 78–130 kHz, which the aggregate score entirely obscures. SUNet achieves lower peak scores but maintains consistent positive recovery across the full 20 kHz to 4 MHz range ($\sigma = 0.54$ vs WaveNet's 2.82), and degrades more gracefully below 20 kHz. The frequency-resolved profile provides a richer and more operationally relevant characterization of both models than the aggregate score alone.

Limitations. SUNet underperforms below 20 kHz due to combined $1/f$ noise dominance and STFT resolution constraints. Peak scores above 20 kHz are lower than WaveNet's. The aggregate score trails the ensemble FCNet baseline.

Future work. Three directions are worth exploring. First, a larger N_{FFT} would increase frequency resolution near DC and could directly address the sub-20 kHz degradation, at the cost of larger spectrograms and higher memory requirements. Second, a 1D convolutional frontend initialized to STFT basis functions and fine-tuned during training could allow the decomposition to adapt to the signal regime; whether this helps in practice would need empirical verification. Third, frequency-aware evaluation metrics could be adopted in future TIDMAD evaluations to better capture per-mass-range denoising behavior.

References

[1] Fry, J. et al. TIDMAD: Time Series Dataset for Axion Dark Matter. *arXiv:2406.04378*, 2024.

[2] Ouellet, J. L. et al. First Results from ABRACADABRA-10cm: A Search for Sub-μeV Axion Dark Matter. *Physical Review Letters*, 122:121802, 2019.

[3] Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI*, 2015.

[4] van den Oord, A. et al. WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499*, 2016.

[5] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal Loss for Dense Object Detection. In *ICCV*, 2017.

[6] Wu, Y. and He, K. Group Normalization. In *ECCV*, 2018.

A Model Architecture and Training Details

A.1 Data Preprocessing: ADC Conversion and Normalization

The TIDMAD HDF5 files store raw SQUID measurements as 8-bit signed integers (int8) over a hardware-determined voltage range. Raw counts are converted to millivolts via:

$$v[n] \;=\; \big(r[n] + 128\big) \cdot \frac{V_{\text{range}}}{256}, \tag{2}$$

where $r[n] \in [-128, 127]$ is the raw int8 sample. Each 1-second segment is then normalized to zero mean and unit variance:

$$\tilde{x}[n] \;=\; \frac{v[n] - \mu}{\sigma + \epsilon}, \quad \epsilon = 10^{-8}, \tag{3}$$

applied per segment and per channel independently. At inference, the model output is denormalized back to millivolt scale ($\hat{v}[n] = \hat{x}[n](\sigma + \epsilon) + \mu$) and converted to int8 for writing to the ABRA HDF5 format expected by the benchmark.

A.2 STFT Frontend and Two-Channel Representation

Let $x \in \mathbb{R}^L$ denote a normalized 1-second segment ($L = 10{,}000{,}000$). The STFT is computed as:

$$X[k,t] \;=\; \sum_{n=0}^{N_{\text{FFT}}-1} x[n + t \cdot H]\, w[n]\, e^{-j2\pi kn/N_{\text{FFT}}}, \tag{4}$$

with Hann window w, $N_{\text{FFT}} = 40{,}000$, hop $H = 20{,}000$, giving $\Delta f = 250$ Hz/bin, $N_F = 20{,}001$ bins, and $T = 501$ frames (`center=True` padding). The complex output is decomposed into a two-channel tensor:

$$\mathbf{X} = \begin{bmatrix} \text{Re}(X[k,t]) \\ \text{Im}(X[k,t]) \end{bmatrix} \in \mathbb{R}^{2 \times N_F \times T}. \tag{5}$$

The axion signal concentrates in bin $k_a = \lfloor f_a/\Delta f \rceil$ as a complex phasor $X_{\text{signal}}[k_a, t] \approx \frac{A}{2} W(0)(\cos\phi + j\sin\phi)$, with negligible energy elsewhere. STFTs are precomputed and cached to disk in BFloat16 precision (286 GB for 4,014 segments), reducing per-epoch wall time from > 6 hours to $\sim$2 hours.

A.3 Architecture

SUNet is a 2D UNet [3] with DoubleConv blocks:

$$\text{DoubleConv}(\mathbf{h}) = \text{GELU}(\text{GN}(\text{Conv}_2(\text{GELU}(\text{GN}(\text{Conv}_1(\mathbf{h}))))), \tag{6}$$

using GroupNorm [6] for stability at batch size $B = 2$. The encoder applies asymmetric pooling:

$$\mathbf{e}^{(1)} = \text{DoubleConv}(\mathbf{X}), \quad \mathbf{e}^{(\ell)} = \text{DoubleConv}(\text{Pool}_\ell(\mathbf{e}^{(\ell-1)})), \quad \ell = 2, 3, 4, \tag{7}$$

Table 3: Encoder tensor shapes for input $[2, 20001, 501]$.

Stage	Shape	Pool applied
Input	$[2,\ 20001,\ 501]$	none
Enc1	$[16,\ 20001,\ 501]$	1×2 before enc2
Enc2	$[32,\ 20001,\ 250]$	2×2 before enc3
Enc3	$[64,\ 10000,\ 125]$	2×2 before enc4
Enc4	$[128,\ 5000,\ 62]$	2×2 before bottleneck
Bottleneck	$[256,\ 2500,\ 31]$	—

with $\mathrm{Pool}_1 = \mathrm{MaxPool}_{1\times 2}$ (time-only) and $\mathrm{Pool}_{\ell\geq 2} = \mathrm{MaxPool}_{2\times 2}$. Time-only pooling at stage 1 preserves all $N_F = 20{,}001$ frequency bins so the network localizes k_a at full resolution before any frequency compression. Channel ladder $C = [16, 32, 64, 128, 256]$; tensor shapes are given in Table 3.

The decoder mirrors the encoder with skip connections:

$$\mathbf{d}^{(\ell)} = \mathrm{DoubleConv}([\mathrm{Upsample}(\mathbf{d}^{(\ell+1)}),\ \mathbf{e}^{(\ell)}]). \tag{8}$$

The output head $\hat{\mathbf{X}} = \mathrm{Conv2d}_{1\times 1}(\mathbf{d}^{(1)}) \in \mathbb{R}^{2\times N_F \times T}$ is recombined into a complex spectrogram and inverted via iSTFT to recover the denoised time series:

$$\hat{x}[n] = \frac{1}{N_{\mathrm{FFT}}} \sum_{t} \sum_{k=0}^{N_F-1} \hat{X}[k,t]\, e^{j2\pi kn/N_{\mathrm{FFT}}} \cdot w[n - tH]. \tag{9}$$

The model minimizes MSE over both STFT channels:

$$\mathcal{L} = \frac{1}{2N_F T} \sum_{c\in\{\mathrm{Re},\mathrm{Im}\}} \sum_{k,t} \big(\hat{X}_c[k,t] - X_c^{\mathrm{clean}}[k,t]\big)^2. \tag{10}$$