

Deep-learning denoising of radio signals for ultra-high-energy cosmic-ray detection

Zhisen Lai^{1,*}, Oscar Macias^{1,2,†}, Aurélien Benoit-Lévy³, Arsène Ferrière^{3,4} and Matías Tueros^{5,6}

¹*Department of Physics and Astronomy, San Francisco State University, San Francisco, CA 94132, USA*

²*GRAPPA—Gravitational and Astroparticle Physics Amsterdam, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

³*Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France*

⁴*Sorbonne Université, CNRS, Laboratoire de Physique Nucléaire et des Hautes Energies (LPNHE), 4 Pl. Jussieu, Paris, 75005, France*

⁵*Instituto de Física La Plata—CCT La Plata - CONICET-UNLP, Diag 113 y 63, La Plata (1900), Argentina*

⁶*Depto. de Física, Fac. de Cs. Ex., Universidad Nacional de La Plata, Casilla de Correo 67, La Plata (1900), Argentina*



(Received 3 February 2026; accepted 8 September 2026; published 28 September 2026)

Extensive air showers initiated by ultra-high-energy cosmic rays produce broadband radio pulses detectable over large areas, but Galactic and instrumental backgrounds limit sensitivity near threshold. The Giant Radio Array for Neutrino Detection (GRAND) will use large arrays of autonomous radio antennas to detect these air showers, requiring reliable recovery of weak pulses from background fluctuations. We develop a convolutional denoiser for GRAND-like antenna traces that combines the time-domain waveform with the magnitude and phase of its Fourier transform. The model is trained on 4.1×10^5 simulated traces constructed from ZHAIRES air-shower emission, a reduced GRAND-like radio-frequency (RF) chain response, and broadband noise normalized to the expected Galactic-noise scale. In these simulations, denoising increases the detection probability at fixed false-positive rate and improves pulse-time recovery near threshold. The reconstructed peak amplitude is attenuated in the noise-dominated regime. This bias decreases as the pulse becomes better resolved, but a smaller channel-dependent bias persists at higher signal-to-noise ratio. Near threshold, denoising therefore primarily improves candidate recovery and timing, while the measured amplitude response provides a basis for calibrating amplitude-sensitive observables.

DOI: [10.1103/29jx-bcgq](https://doi.org/10.1103/29jx-bcgq)

I. INTRODUCTION

The flux of ultra-high-energy cosmic rays (UHECRs) is low at the highest energies, requiring large exposure for their detection [1,2]. The Giant Radio Array for Neutrino Detection (GRAND) is a planned radio observatory for ultra-high-energy neutrinos, cosmic rays, and gamma rays [3]. GRAND will detect these particles through the coherent radio emission produced by the extensive air showers they generate, with sensitivity concentrated in the 50–200 MHz band. Near threshold, these short radio pulses must be identified against ambient and instrumental backgrounds.

Galactic background radiation and human-made radio-frequency interference (RFI) dominate the noise floor at most sites and can mask faint cosmic-ray pulses [4,5]. These backgrounds raise the effective detection threshold and reduce sensitivity to weak air showers. Radio detection near threshold therefore requires recovery of short cosmic-ray pulses from contaminated waveforms, with improved effective signal-to-noise ratio (SNR) and a lower detection threshold.

Classical filtering methods have been investigated for this purpose. In particular, matched filtering [6] using an expected pulse template can improve sensitivity under idealized noise conditions, such as white noise, and has been shown in simulations to reduce trigger thresholds. However, its performance degrades under realistic non-Gaussian noise because RFI does not satisfy the assumptions of an idealized random background.

Wavelet methods isolate transient structure on several scales and have been applied to narrow-band interference and Gaussian noise [7]. Early digital radio detectors also used iterative subtraction based on the radio-astronomy CLEAN algorithm [8] together with beamforming [9]. The

*Contact author: zlai@sfsu.edu

†Contact author: macias@sfsu.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

performance of these approaches depends on the assumed signal and background properties, particularly for complex or nonstationary interference. Analyses therefore often impose high post-filtering SNR thresholds that reject weak air showers [10].

Learned denoising methods have been applied to signal-reconstruction problems in radio astronomy and gravitational-wave data analysis [11–14]. For air-shower radio data, deep-learning methods have been used to identify weak pulses and reconstruct noise-free waveforms [15–18]. In simulated data, Ref. [15] reported a pulse-energy resolution of order 20% without bias and suppression of narrow-band RFI. Tunka-Rex Collaboration found that an autoencoder lowered the detection threshold in measured data while preserving shower reconstruction [16]. An IceCube analysis of several months of data recovered about five times more events than a conventional high-threshold trigger, with improved purity and unchanged angular resolution [18]. For GRAND-like traces, however, timing, waveform fidelity, and amplitude response must be validated separately.

We extend the initial study of Ref. [19] using a denoising autoencoder that processes the three polarization traces jointly. The encoder combines local time-domain features with the magnitude and phase of the Fourier transform. During training, each contaminated detector-level trace is paired with its noise-free reference.

We evaluate the model using antenna-level simulations that include a reduced GRAND-like radio-frequency (RF) chain response and broadband stochastic noise normalized to the expected Galactic-noise scale. These simulations omit the full local sidereal time (LST) dependence and spectral structure of the sky background, nonstationary anthropogenic RFI, and variations in the detector response. Dedicated GRAND studies, including NUTRIG [20,21], address self-triggering and background rejection under more realistic operating conditions. We restrict the present study to antenna-level denoising under these controlled conditions. Incorporating the full detector response and more realistic backgrounds is left to future studies.

We demonstrate that our proposed denoiser increases the detection probability at fixed false-positive rate and improves pulse-time recovery near threshold for this detector and noise model. At low SNR, however, the network attenuates part of the weak pulse, and a smaller channel-dependent bias remains at higher SNR. We therefore characterize timing, amplitude response, and waveform fidelity separately, since denoising affects these observables differently near threshold.

The remainder of this paper is organized as follows. Section II describes the simulated data and preprocessing, and Sec. III presents the denoiser architecture, training procedure, and performance metrics. Section IV reports the results. We discuss their implications and limitations in Sec. V and summarize our conclusions in Sec. VI.

II. TRAINING DATA AND PREPROCESSING

A. Air-shower and radio-emission simulations

We simulate the intrinsic radio emission of extensive air showers with ZHAIRES [22,23]. The sample contains downward-going proton- and iron-induced showers with energies 0.4–4 EeV, zenith angles $37^\circ \leq \theta \leq 87^\circ$, and uniform azimuthal coverage. We adopt the geomagnetic field and atmospheric profile of *GRANDProto300* [24]. The ZHAIRES electric-field waveform defines the incident signal $v_{\text{sig}}(t)$ used to construct the detector-level reference and noisy traces.

B. Galactic and instrumental noise model

A GRAND-like detection unit passes the antenna voltage through an analog RF chain containing amplifiers, cables, filters, and digitization electronics [25]. The net transfer function modifies the pulse spectrum, defines the analysis band, and introduces a propagation delay. Diffuse Galactic emission contributes to the antenna voltage noise [25].

For the training sample, we replace the full component-level implementation in *GRANDLib* [25] with a reduced RF-chain model. A fixed phenomenological transfer function approximates the net gain and band limitation of the low-noise amplifier (LNA), cable, and filter chain. Each trace also includes an independent timing offset of a few nanoseconds and broadband Gaussian noise normalized to the expected Galactic-noise scale (cf. Fig. 1). The model excludes the LST-dependent colored sky background, measured RFI, and time-dependent detector conditions.

For an incident signal $v_{\text{sig}}(t)$, the noise-free reference and contaminated input are

$$x_{\text{true}}(t) = \mathcal{R}\{v_{\text{sig}}(t - \delta t)\}, \quad (1)$$

$$x_{\text{noisy}}(t) = \mathcal{R}\{v_{\text{sig}}(t - \delta t) + n(t)\}. \quad (2)$$

Here, $\mathcal{R}$ denotes the reduced RF-chain response, δt the per-trace timing offset, and $n(t)$ the additive broadband-noise realization. The first expression defines the training target, and the second defines the network input. The resulting sample tests antenna-level denoising and is not a full GRAND detector simulation.

C. Dataset construction

The dataset contains 410,673 three-polarization traces (X, Y, Z), each stored with 1024 samples per channel at a sampling interval of 0.5 ns. The network is trained on 512-sample windows extracted from these traces. During training, random pulse-position augmentation prevents the network from exploiting a fixed pulse location. The Ray Tune training procedure partitions the data into training, validation, and unused remainder subsets within each trial,

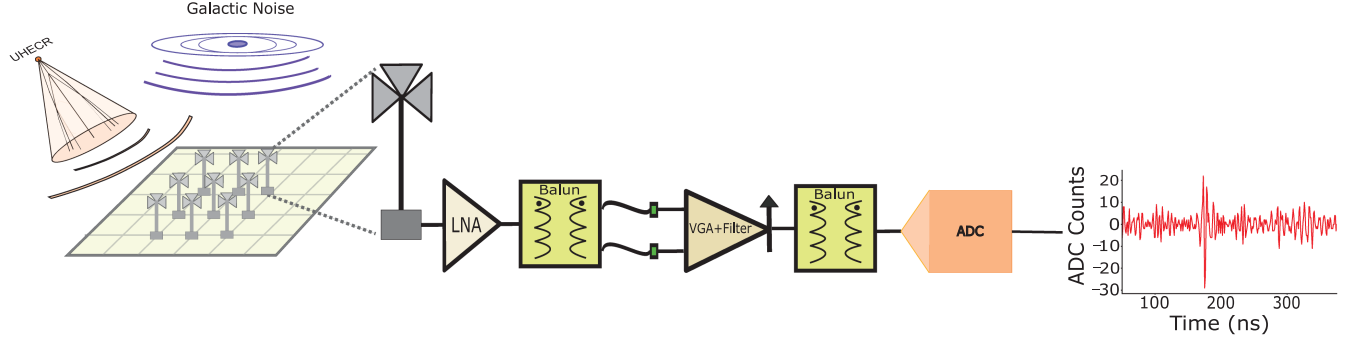


FIG. 1. *Construction of the simulated traces.* A UHECR air shower produces a short radio pulse at a GRAND HORIZON-like antenna [3]. The simulated pulse is passed through a simplified GRAND-like RF-chain model that imposes the band limitation and net gain used in this antenna-level study. This gives the noise-free reference trace. For the network input, broadband stochastic noise is added before applying the same response. These processed traces define the detector-level waveforms used throughout the analysis.

as described in Sec. III D. The network maps $x_{\text{noisy}}(t)$ to the reconstruction $x_{\text{rec}}(t)$ and is trained against $x_{\text{true}}(t)$.

Figure 2 shows the input-SNR distribution of the evaluation sample. Most traces lie near threshold, where noise can displace the reconstructed pulse time and distort amplitude-sensitive quantities. We report timing recovery, peak-amplitude bias, and waveform fidelity separately because these quantities respond differently to denoising. At high SNR, the pulse is already resolved, so denoising has a smaller effect.

III. DENOISER ARCHITECTURE AND TRAINING

A. Denoiser architecture

Figure 3 shows the denoiser architecture. The network processes the three polarization traces jointly through time-

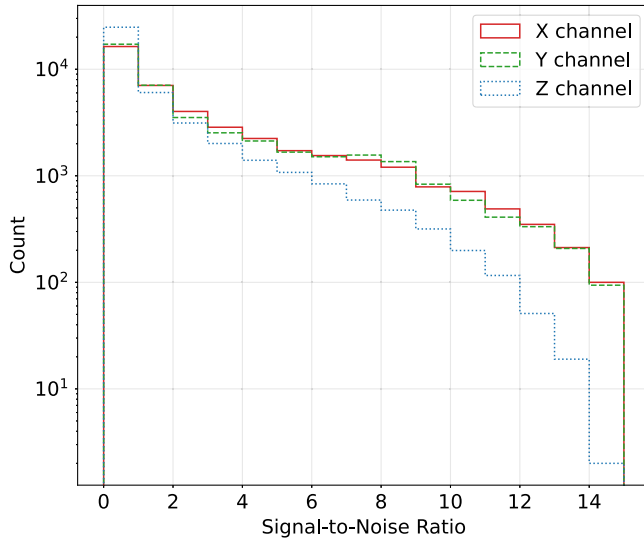


FIG. 2. *Input-SNR distribution of the evaluation sample.* Histogram of SNR_{in} for the simulated GRAND-like antenna traces in the three polarization channels (X, Y, Z). The input SNR is defined in Eq. (10).

domain and Fourier-domain pathways [26]. The time-domain pathway extracts local pulse structure, while the Fourier-domain pathway receives the magnitude and principal phase of the one-sided real Fourier transform of the 512-sample input window. No taper is applied before the transform.

Each encoder branch contains a stem one-dimensional convolution, two residual blocks [27], and two max-pooling operations, giving a net downsampling factor of four. The time, magnitude, and phase branches each end with 256 channels. Concatenating their outputs produces 768 channels, which the fusion module reduces to 384. A three-stage transposed-convolution decoder with effective widths (384, 64, 32, 3) reconstructs the three polarization waveforms. Table I lists the selected configuration, and Appendix B gives further implementation details. An earlier public release of the code is archived on Zenodo (DOI: <https://doi.org/10.5281/zenodo.18233878>) [28]. The implementation and analysis products used here are available in the public `raytune_lib_final_v1` branch [29].

B. Training objective

We train the fiducial model with a multidomain L_1 objective. Its time-domain term is

$$L_{\text{time}} = \frac{1}{N} \sum_t |x_{\text{true}}(t) - x_{\text{rec}}(t)|. \quad (3)$$

The one-sided real Fourier transforms of the reference and reconstructed traces are

$$\tilde{x}_{\text{true}}(f) = \text{RFFT}[x_{\text{true}}(t)], \quad \tilde{x}_{\text{rec}}(f) = \text{RFFT}[x_{\text{rec}}(t)]. \quad (4)$$

The magnitude term is

$$L_{\text{mag}} = \frac{1}{N_f} \sum_f ||\tilde{x}_{\text{true}}(f)| - |\tilde{x}_{\text{rec}}(f)||. \quad (5)$$

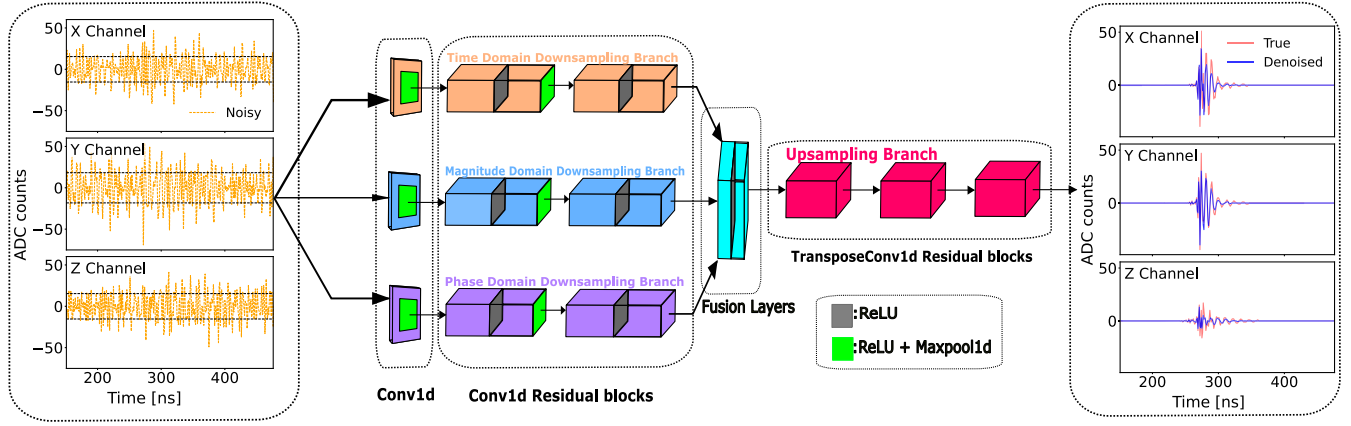


FIG. 3. *Encoder-decoder architecture.* The input contains three polarization channels with 512 samples per channel, extracted from simulated traces stored with 1024 samples. The encoder contains a time-domain branch (orange) and a Fourier-domain branch that separates magnitude (dark blue) and phase (purple). Each branch begins with a one-dimensional convolution followed by two residual blocks. The fusion module (cyan) combines the three feature maps using two additional convolutional layers. The decoder (red) returns denoised traces with the same dimensions as the input.

The fiducial loss uses the direct difference between the principal Fourier phases. With $\phi_{\text{true}} = \arg \tilde{x}_{\text{true}}$ and $\phi_{\text{rec}} = \arg \tilde{x}_{\text{rec}}$, the phase term is

$$L_{\text{phase}} = \frac{1}{N_f} \sum_f |\phi_{\text{rec}}(f) - \phi_{\text{true}}(f)|. \quad (6)$$

The complete objective is

$$L_{\text{multi}} = L_{\text{time}} + \sigma_{\text{mag}} L_{\text{mag}} + \sigma_{\text{phase}} L_{\text{phase}}. \quad (7)$$

The weights $\sigma_{\text{mag}} = 0.011676$ and $\sigma_{\text{phase}} = 0.028893$ were selected on the validation set. The direct principal-phase difference is discontinuous at the $-\pi/\pi$ boundary. To quantify the effect of this discontinuity, we evaluated a wrapped alternative on a separate fixed sample of

41,067 traces by replacing the phase residual with $\text{atan2}[\sin(\Delta\phi), \cos(\Delta\phi)]$ while leaving the other terms unchanged. The mean phase term changed from 2.0711 to 1.5616, whereas the complete objective changed by only 0.0043%. The reported analyses therefore use the checkpoint trained with the direct principal-phase convention.

C. Performance metrics

We evaluate waveform error, pulse timing, and amplitude response with separate metrics. The per-trace mean-squared error is

$$\text{MSE} = \frac{1}{N} \sum_t [x_{\text{true}}(t) - x_{\text{rec}}(t)]^2. \quad (8)$$

TABLE I. Selected configuration of the fiducial multidomain denoiser. The effective decoder widths account for replacement of the unused first nominal decoder entry by the fusion width.

Parameter	Selected value
Objective	Time + magnitude + direct principal-phase L_1
Input window	3×512 samples
Time stem channels	64
Time residual channels	(128, 256)
Magnitude/phase stem channels	64
Magnitude/phase residual channels	(128, 256)
Fusion width	384
Effective decoder widths	(384, 64, 32, 3)
Magnitude-loss weight	0.011676
Phase-loss weight	0.028893
Batch size	1024
Base/maximum learning rate	$1.656 \times 10^{-5} / 5.067 \times 10^{-4}$
Cyclic step size and mode	10,000; triangular2
Weight decay	2.566×10^{-4}
Completed/maximum epochs	63/100
Ray Tune trials	36

The corresponding peak signal-to-noise ratio is

$$\text{PSNR} = 10 \log_{10} \left(\frac{A_{\text{true}}^2}{\text{MSE}} \right), \quad A_{\text{true}} = \max_t |\mathcal{H}[x_{\text{true}}(t)]|, \quad (9)$$

where $\mathcal{H}$ denotes the analytic-signal operation used to compute the Hilbert-envelope magnitude.

We bin the diagnostics by the simulation-level ratio of the clean-trace peak to the full-trace standard deviation of the noisy input,

$$\text{SNR}_{\text{in}} = \frac{P_{\text{true}}}{\sigma_{\text{full}}}, \quad P_{\text{true}} = \max_t x_{\text{true}}(t), \quad \sigma_{\text{full}} = \text{std}_t [x_{\text{noisy}}(t)]. \quad (10)$$

Both P_{true} and σ_{full} are evaluated over the complete 512-sample trace. The denominator therefore includes the signal interval and is not an off-pulse estimate of the noise level. We use this convention throughout the analysis. The quantity serves only to order and bin the simulated traces and should not be compared directly with SNR values based on a separately defined noise window.

The peak-amplitude and timing diagnostics use the analytic-signal envelope over the 512-sample window. Its global maximum defines the peak amplitude and sample index for the clean, noisy, and denoised traces. We apply no additional waveform denoising to the noisy-input baseline. Appendix A describes the separate band-limited waveform diagnostic.

D. Optimization, tuning, and data splits

We train the denoiser with the noisy and clean traces described in Sec. II. In the original Ray Tune search, each trial generated its own random 80/10/10 partition of the 410,673 traces. The training and validation subsets were used to fit the network and assess that trial, while the remaining 10% did not enter the trial objective. The indices for these historical trial-specific partitions were not retained, so the exact partition used by the selected trial cannot now be reconstructed.

For the revised timing, peak-amplitude, and waveform-fidelity diagnostics, we use a fixed evaluation subset generated with seed 12345. It contains 41,068 traces and ensures that all revised diagnostics are evaluated on the same reproducible sample. Because the overlap between this subset and the historical training partition of the selected trial cannot be established, we refer to it as an evaluation subset rather than as an independent held-out test set.

The network is implemented in PyTorch [30] and optimized with Adam and weight decay [31]. The learning rate follows a triangular2 cyclical schedule [32], with the minimum and maximum rates and cycle length listed in Table I.

We explore the hyperparameter space with Ray Tune [33] and use the Asynchronous Successive Halving

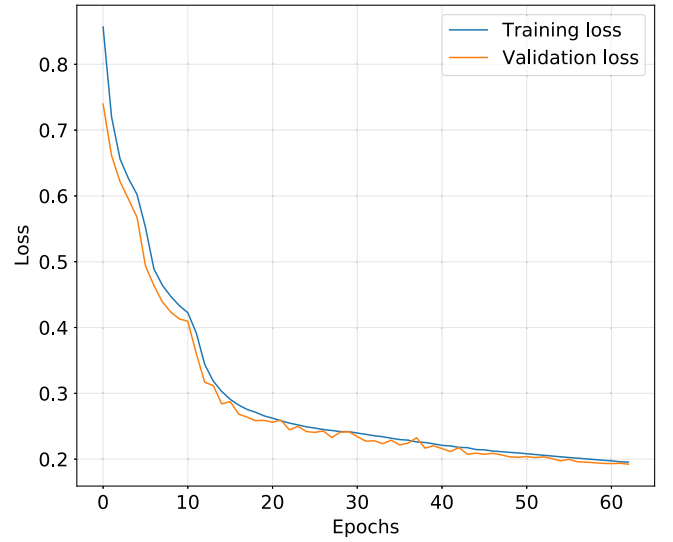


FIG. 4. Training and validation values of the complete multi-domain objective in Eq. (7) for the selected trial. Each constituent term is averaged over the batch, channels, and its time or frequency bins before the weighted sum is formed. The selected Asynchronous Successive Halving Algorithm (ASHA) trial completed 63 of the maximum 100 epochs.

Algorithm (ASHA) [34] to distribute the training budget among the 36 trials. ASHA uses the validation loss to decide which configurations should receive additional training, so the maximum of 100 epochs is an upper budget for a trial rather than a prescribed training length. Under this procedure, the selected trial completed 63 epochs. Figure 4 shows its training and validation histories, Table I gives the selected configuration, and Table II lists the ranges explored in the search.

As a separate architecture check, Appendix D compares the dual-branch network with a time-only model. For this comparison, both models are retrained from scratch using the same seeded data manifest and training configuration, and are evaluated on the same untouched test subset of 41,068 traces. The dual-branch model reaches a mean test PSNR of 33.738 dB, compared with 33.341 dB for the time-only model; the corresponding multidomain losses are 16.314 and 17.574. Appendix D defines the loss and PSNR conventions used specifically for this ablation.

IV. RESULTS

A useful denoiser must do more than produce visually cleaner traces. Near threshold, it should increase the probability of recovering a weak pulse without increasing the false-positive rate, preserve the pulse arrival time used in wave front reconstruction, and retain the waveform information relevant to amplitude-sensitive observables (e.g., Refs. [35,36]). These requirements probe different aspects of the reconstruction. A pulse can be recovered at approximately the correct time even when its amplitude or

TABLE II. Hyperparameter search ranges for the dual-branch denoiser.

Parameter	Search space
Time-branch stem channels	{64, 128}
Time-branch residual channels	{(64, 128), (128, 256)}
Frequency-branch stem channels	{64, 128}
Frequency-branch residual channels	{(64, 128), (128, 256)}
Nominal decoder channels ^a	{(128, 64, 32, 3), (256, 128, 64, 3), (64, 32, 16, 3), (32, 16, 8, 3)}
Base learning rate	Log-uniform in $[10^{-6}, 10^{-4}]$
Maximum learning rate	Log-uniform in $[10^{-5}, 10^{-3}]$
Cyclic step size	{10, 000, 20, 000, 30, 000, 40, 000, 50, 000}
Learning-rate mode	{triangular, triangular2}
Batch size	{512, 1024, 2048, 4096}
Weight decay	Log-uniform in $[10^{-5}, 5 \times 10^{-3}]$
Magnitude-loss weight	Log-uniform in $[10^{-2}, 1]$
Phase-loss weight	Log-uniform in $[10^{-2}, 1]$

^aThe fiducial model replaces the first nominal decoder entry with the fusion width. Only the remaining decoder entries affect the instantiated model.

detailed waveform is imperfect. We therefore characterize three complementary aspects of the reconstruction: detection at fixed false-positive rate, pulse timing and peak-amplitude response, and band-limited waveform fidelity. The latter is discussed separately in Appendix A.

Figure 5 shows representative traces from the fiducial denoiser in the three polarization channels. At $\text{SNR} \simeq 4.2$, the reconstructed waveform closely follows the reference pulse while suppressing incoherent baseline fluctuations. At $\text{SNR} \simeq 2.5$, where the pulse is less distinct in the noisy trace, the network recovers a compact feature near the injected pulse time and suppresses much of the broadband noise, although the reconstructed peak amplitude is attenuated. The frequency-domain panels show the corresponding reduction in stochastic spectral power. The ensemble tests below quantify the same behavior. Near threshold, the largest gains occur in pulse recovery and timing, whereas the amplitude response requires separate evaluation.

A. Timing recovery and amplitude response

Timing and amplitude can respond differently to denoising. For direction reconstruction, the relevant quantity is the pulse arrival time at each antenna, whereas amplitude-sensitive observables depend on the fidelity of the reconstructed electric-field scale (e.g., Refs. [35–37]). Near threshold, these requirements separate because the denoiser can recover the arrival time of a weak pulse even when its reconstructed amplitude remains imperfect.

We first examine pulse timing. To compare the noisy and denoised traces for the same population of pulse candidates, we apply a common triggerlike preselection based only on the noisy waveform. Let $e_{\text{noisy}} = |\mathcal{H}[x_{\text{noisy}}]|$ and let A_{noisy} denote its maximum. We estimate the local envelope scale from the median absolute deviation after excluding 32 samples on either side of this maximum,

$$\sigma_{\text{env}} = 1.4826 \text{ median}\{|e_{\text{noisy}} - \text{median}(e_{\text{noisy}})|\}, \quad (11)$$

where the medians are evaluated over the remaining samples. We retain traces satisfying $A_{\text{noisy}} \geq 2\sigma_{\text{env}}$. This selection only defines a common sample for the timing comparison.

We regard the pulse time as recovered when the maximum of the denoised Hilbert envelope lies within ten samples of the clean reference, corresponding to 5 ns at the 0.5 ns sampling interval. Figure 6 shows that denoising reduces large timing errors at low and intermediate SNR. The improvement decreases as the pulse emerges above the background. Once the signal dominates the trace, the noisy waveform already identifies the peak time reliably, and the timing efficiency approaches unity.

The amplitude response behaves differently. We do not apply the triggerlike timing selection for this comparison. Instead, we compare the noisy and denoised peak amplitudes for the same trace-channel pairs in the fixed evaluation subset and bin them by the common SNR_{in} . Figure 7 shows two competing effects near threshold. In the noisy trace, the largest envelope excursion is often set by a positive noise fluctuation, biasing the measured peak upward. Denoising suppresses these fluctuations, but for the weakest pulses it also removes part of the signal and shifts the reconstructed peak amplitude low. This attenuation decreases with increasing SNR, although a smaller channel-dependent bias remains over part of the displayed range.

The two diagnostics therefore separate near threshold. Denoising improves pulse-time recovery in a regime where the raw waveform often fails to localize the pulse, while the reconstructed amplitude retains an SNR- and channel-dependent response. Translating the timing gain into improved shower-direction reconstruction will depend

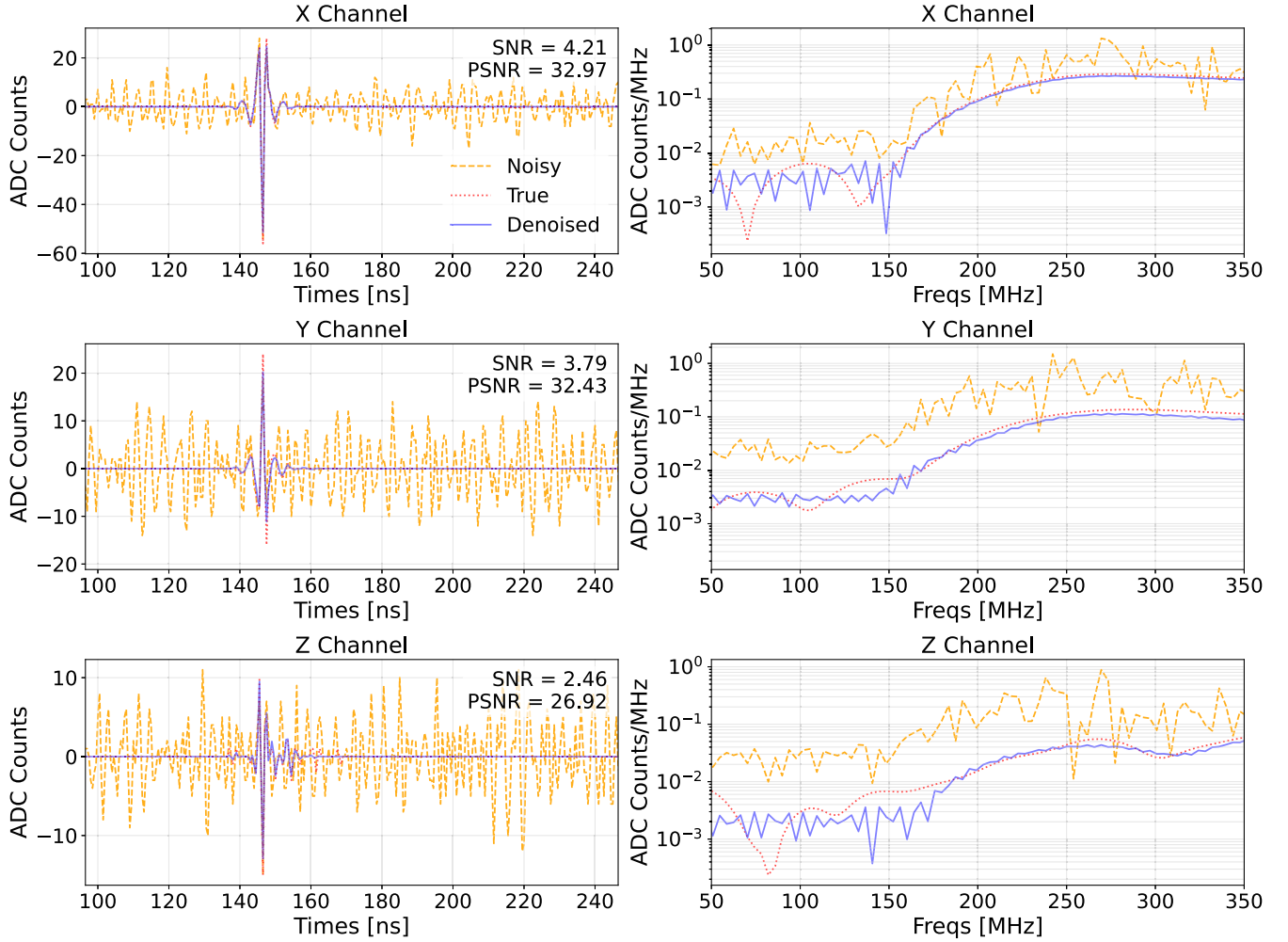


FIG. 5. *Representative reconstructions in the time and frequency domains.* The rows show different antenna traces in the X, Y, and Z channels, with high, intermediate, and low input SNR, respectively. The left panels show the clean target (red dotted), noisy input (orange dashed), and denoised reconstruction (blue solid) in the time domain. The right panels show the corresponding Fourier-amplitude spectra. The time and frequency axes use the simulation sampling interval, $\Delta t = 0.5$ ns. The annotations give the per-channel input SNR defined in Eq. (10) and the peak signal-to-noise ratio (PSNR) defined in Eq. (9).

on which antennas are recovered, their spatial configuration, and the distribution of timing residuals across the event.

B. Detection at fixed false-positive rate

To estimate detection probability at a fixed false-positive rate, we construct signal-free traces from off-pulse windows. For each event, we locate the clean pulse, mask a conservative interval around it, and draw segments from the remaining contaminated trace to assemble a full-length noise-only waveform. A candidate identified in this sample is counted as a false positive [38].

We calibrate separate candidate thresholds for the noisy input and denoised output using their respective noise-only samples. Each threshold is chosen to give one false positive per 1000 traces, $P_{\text{FP}} = 10^{-3}$, and is then applied to signal-containing traces. Figure 8 shows the resulting detection probability as a function of injected SNR. At the same

false-positive rate, the denoiser recovers more signal-containing traces, with the largest difference near threshold.

V. DISCUSSION

The clearest effect of denoising is the recovery of pulse timing near threshold. Planar wave front reconstruction depends on the relative arrival times of the radio pulse across the array and on the antenna positions [35]. Recovering a reliable timestamp from a trace in which the pulse is otherwise difficult to localize can therefore contribute directly to the wave front fit. The resulting improvement in shower-direction reconstruction will depend on the spatial configuration of the recovered antennas and the distribution of their timing residuals, rather than simply on the number of recovered antennas.

The amplitude response exhibits a different behavior near threshold. In the noisy trace, maximizing the envelope tends to select positive noise fluctuations and can bias the

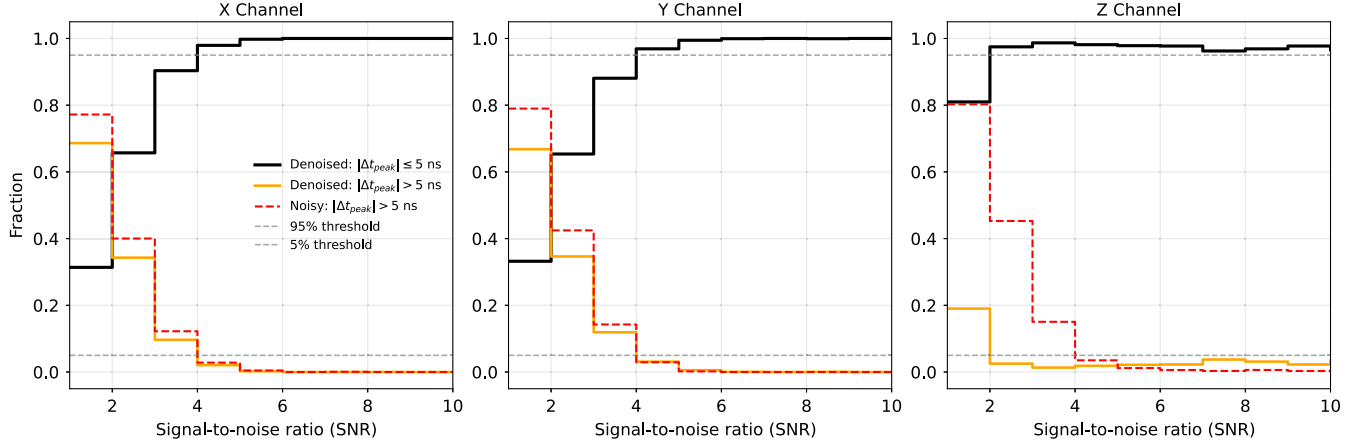


FIG. 6. *Timing reconstruction versus input SNR.* For each polarization channel, the black curve shows the fraction of traces passing the noisy-input preselection for which the denoised peak time agrees with the clean reference within ten samples, or 5 ns at the 0.5 ns sampling interval. The orange curve shows the complementary denoised failure fraction, $|\Delta t_{\text{peak}}| > 5$ ns, and the red dashed curve shows the corresponding failure fraction for the noisy input. Gray dashed lines mark the 95% and 5% levels.

measured peak upward. The denoiser removes much of this stochastic contribution, but for the weakest pulses it also suppresses part of the signal, producing the negative amplitude bias seen in Fig. 7. This attenuation decreases as the pulse becomes better resolved, although a smaller channel-dependent response remains at higher SNR. Timing and amplitude therefore separate in the noise-dominated regime. In particular, the pulse arrival time can be recovered even when its amplitude remains biased.

This distinction is particularly relevant for energy-sensitive radio observables. The measured electric field is a superposition of signal and noise, and random relative phases can produce constructive or destructive interference at low SNR [10]. Amplitude and fluence reconstruction therefore require explicit treatment of the noise contribution even without denoising. Here, the network introduces an additional amplitude response that we measure as a

function of SNR and polarization. Characterizing this response is a first step toward calibrating amplitude-sensitive observables.

The fixed-false-positive comparison gives a complementary view of the near-threshold gain. At $P_{\text{FP}} = 10^{-3}$, the denoised traces have a higher detection probability than the noisy traces, with the largest difference where the pulse is only weakly resolved. The improvement in timing is therefore accompanied by better separation of signal-containing and noise-only traces under the simulated broadband background considered here.

The matched time-only comparison in Appendix D provides a separate check of the role of the Fourier pathways. Under the same training configuration, the dual-branch model improves the mean test PSNR by 0.40 dB and gives a lower multidomain loss than the time-only model. The effect is modest, but both metrics favor the dual-branch

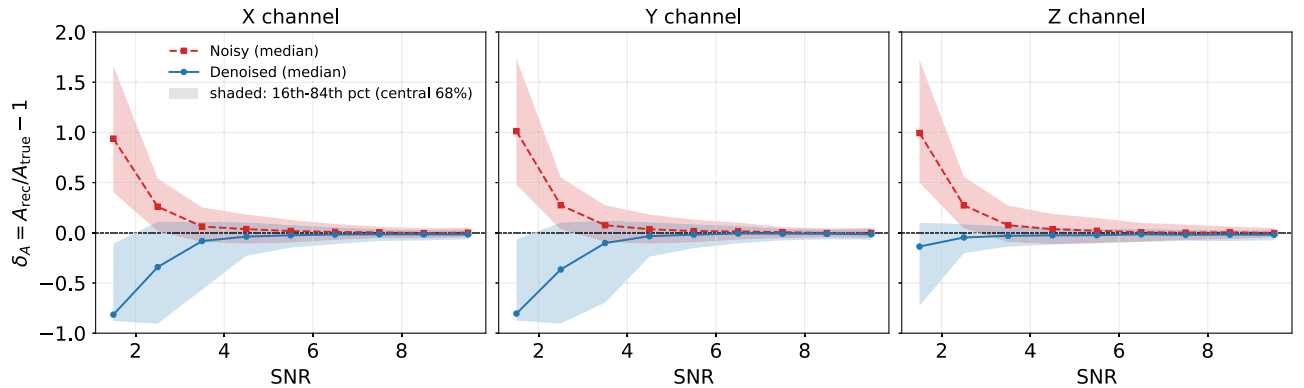


FIG. 7. *Peak-amplitude bias versus input SNR.* Median signed bias, $\delta_A = A_{\text{rec}}/A_{\text{true}} - 1$, binned in input SNR for trace-channel pairs from the fixed evaluation subset. The displayed unit-width bins span $1 < \text{SNR}_{\text{in}} < 10$; bins containing fewer than 20 trace-channel pairs are omitted. The noisy and denoised curves use identical trace-channel pairs and SNR values, with no triggerlike selection applied. Shaded bands show the central 68% population interval (16th–84th percentiles). At low SNR, the noisy peak is biased high by maximization over noise fluctuations, whereas the denoised peak is biased low. The attenuation decreases as the pulse becomes better resolved, although a residual channel-dependent bias remains.

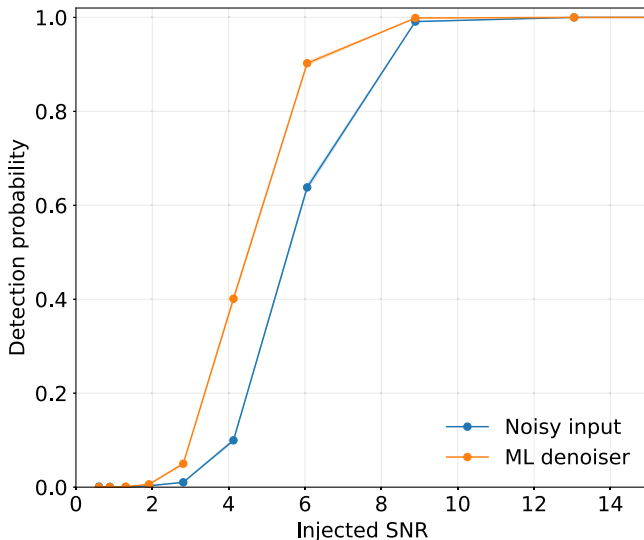


FIG. 8. *Detection probability versus injected SNR at fixed false-positive rate, $P_{\text{FP}} = 10^{-3}$.* The blue and orange curves show candidate selection on the noisy traces and on the traces denoised using machine learning (ML), respectively. For each method, the threshold is calibrated using its own constructed noise-only sample to enforce the same false-positive rate. The comparison therefore measures the gain in detection probability under the simulated broadband-noise background considered here.

architecture. We therefore view the frequency-domain pathways as a useful component of the denoiser rather than the main source of its performance.

VI. CONCLUSIONS

We developed a convolutional denoiser for GRAND-like radio traces that combines time-domain information with Fourier-domain magnitude and phase. In the simulations considered here, denoising increases the detection probability at fixed false-positive rate and improves pulse-time recovery near threshold. In particular, the network recovers timing information from traces in which the pulse is otherwise difficult to localize. Because relative pulse times are the basic input to radio wave front reconstruction, this recovery increases the timing information available for air-shower reconstruction.

The amplitude response behaves differently. For weak pulses, denoising suppresses stochastic fluctuations but also attenuates part of the signal, producing a negative peak-amplitude bias. This attenuation decreases as the pulse becomes better resolved, while a smaller channel-dependent response persists at higher SNR. Together with the band-limited waveform test, these results show that pulse timing and overall waveform recovery can improve before the peak amplitude reaches comparable fidelity. The measured SNR- and polarization-dependent amplitude response provides a basis for calibrating amplitude-sensitive observables.

For the present denoiser, candidate recovery and timing show the largest near-threshold improvements, while the amplitude response must be accounted for in energy-sensitive reconstruction. The next step is to test these gains with more complete GRAND detector simulations and measured background conditions, and to propagate the recovered pulse information through shower direction and energy reconstruction.

ACKNOWLEDGMENTS

We thank Sara El Bouch, Claire Guépin, and Simon Prunet for useful discussions. We also thank GRAND Collaboration for providing the *GRANDProto300* site parameters and related technical information used in this study. O. M. and Z. L. acknowledge support from the U. S. National Science Foundation under Grant No. 2418730.

DATA AVAILABILITY

The code repository is available on GitHub [29]. The analysis scripts, evaluation manifests, figure-provenance records, and figures used in this study are available in the public `raytune_lib_final_v1` branch [29]. The earlier public release is archived on Zenodo (DOI: <https://doi.org/10.5281/zenodo.18233878>). The clean waveforms are generated with ZHAIRES. Section II describes the detector-response and noise model. Additional data products are available from the authors upon reasonable request.

APPENDIX A: BAND-LIMITED WAVEFORM-FIDELITY DIAGNOSTIC

We test whether identical band conditioning leaves the denoised output closer to the injected waveform than the noisy input. The fiducial model is evaluated on a fixed seed-12345 evaluation subset of 41,068 traces. Because the historical training partition of the selected Ray Tune trial was not archived, we do not assume that this evaluation subset is disjoint from that trial's training data. This is a waveform-reconstruction diagnostic, not a calibration of amplitude, fluence, or primary energy.

We use the sampling interval of the simulated traces, $\Delta t = 0.5$ ns, and apply a fourth-order Butterworth band-pass over $\mathcal{B} = [50, 200]$ MHz to the reference, noisy, and reconstructed traces [25]. The nominal fidelity interval is centered on the clean Hilbert-envelope maximum and extends by 150 ns on either side. At this sampling interval, the nominal interval spans 601 samples, which is wider than the 512-sample evaluation trace. The interval is therefore clipped at the trace boundaries, and the fidelity metrics are evaluated over the full 512-sample trace. We denote this effective interval by $\mathcal{W}$ below. For each polarization channel, the normalized mean-squared error (NMSE) is

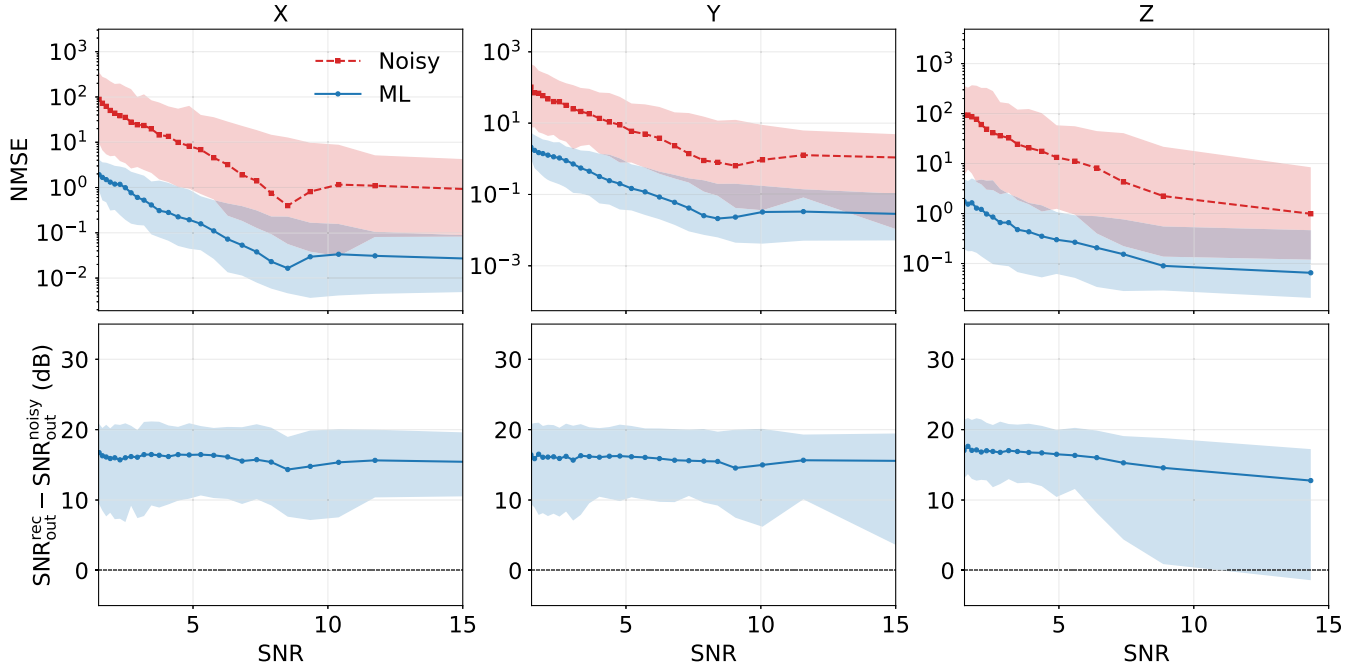


FIG. 9. *Band-limited waveform fidelity.* The top row shows the rolling median NMSE_B for the ML reconstruction and the bandpassed noisy-input baseline after applying the fourth-order 50–200 MHz Butterworth bandpass. The bottom row shows $\Delta\text{SNR}_{\text{out}} = \text{SNR}_{\text{out}}(x_{\text{rec}}^B) - \text{SNR}_{\text{out}}(x_{\text{noisy}}^B)$. Both quantities are evaluated over the full 512-sample, band-limited trace and binned using the full-trace input SNR in Eq. (10). The nominal ± 150 ns fidelity interval is wider than the trace at $\Delta t = 0.5$ ns and is therefore clipped to the available samples. Shaded bands contain the central 90% population interval in the rolling windows. A truth-dependent clean-power gate removes the lowest 10% of positive clean powers in each channel. Lower NMSE or higher output SNR indicates closer band-limited agreement with the reference waveform but does not establish an unbiased amplitude or fluence estimator.

$$\text{NMSE}_B = \frac{\sum_{t \in \mathcal{W}} [x_{\text{rec}}^B(t) - x_{\text{true}}^B(t)]^2}{\sum_{t \in \mathcal{W}} [x_{\text{true}}^B(t)]^2}. \quad (\text{A1})$$

The noisy-input baseline is obtained by replacing x_{rec}^B with the bandpassed noisy trace over the same interval $\mathcal{W}$.

For a candidate waveform $y \in \{x_{\text{noisy}}^B, x_{\text{rec}}^B\}$, we define

$$\text{SNR}_{\text{out}}(y) = 10 \log_{10} \left[\frac{\sum_{t \in \mathcal{W}} (x_{\text{true}}^B(t))^2}{\sum_{t \in \mathcal{W}} (y(t) - x_{\text{true}}^B(t))^2} \right]. \quad (\text{A2})$$

The lower panels of Fig. 9 show $\Delta\text{SNR}_{\text{out}} = \text{SNR}_{\text{out}}(x_{\text{rec}}^B) - \text{SNR}_{\text{out}}(x_{\text{noisy}}^B)$. To avoid ill-conditioned ratios in channels with negligible injected power, we exclude trace-channel pairs below the tenth percentile of positive clean power in $\mathcal{W}$, separately for each channel. This truth-dependent gate is applied identically to the noisy and denoised curves and is used only for this simulation diagnostic.

Figure 9 shows lower band-limited NMSE and a positive output-SNR difference for the ML reconstruction over much of the displayed range. The largest change occurs in the noise-dominated regime. As Fig. 7 shows, however, a lower integrated waveform error can coexist with a biased peak amplitude.

APPENDIX B: AUTOENCODER ARCHITECTURE

The network maps a noisy window $x_{\text{noisy}} \in \mathbb{R}^{3 \times 512}$ to a reconstruction x_{rec} of the same shape. The time-domain pathway begins with a one-dimensional convolution with 64 output channels, followed by residual stages with 128 and 256 channels. The magnitude and phase pathways use the same channel progression. Two max-pooling operations in each branch give a downsampling factor of 4.

The final 256-channel outputs of the time, magnitude, and phase branches are concatenated to form 768 channels. The fusion module reduces this representation to 384 channels. The effective decoder widths are (384, 64, 32, 3). The fusion width overwrites the first nominal decoder entry in the Ray Tune configuration, so this entry is not an independent model parameter. Transposed convolutions restore the 512-sample length and return the three reconstructed polarization traces.

APPENDIX C: TRAINING AND VALIDATION

We implement the model in PyTorch [30] and optimize it with Adam and weight decay [31]. The learning rate follows a triangular2 cyclic schedule [32]. We select hyperparameters with Ray Tune [33] and ASHA [34] over 36 trials with a maximum budget of 100 epochs. The

selected trial completed 63 epochs. Each historical Ray Tune trial generated its own random training and validation partitions; the exact indices used by the selected trial were not archived, as discussed in Sec. III D.

APPENDIX D: TIME-ONLY COMPARISON OF THE DENOISER ARCHITECTURE

We compare the dual-branch model with a time-only model under the same training configuration. The two models use the same 512-sample inputs, seeded 80/10/10 data manifest, preprocessing, pulse-position augmentation, channel widths, fusion implementation, optimizer, cyclic learning-rate schedule, batch size, and 100-epoch training budget. They differ only in whether the Fourier-magnitude and phase pathways are enabled. No additional hyperparameter search is performed.

For this ablation, we use the selected channel widths and optimizer hyperparameters and retrain both models from scratch. Both models are trained with the wrapped angular phase residual

$$\Delta\phi_{\text{wrap}} = \text{atan2}[\sin(\phi_{\text{rec}} - \phi_{\text{true}}), \cos(\phi_{\text{rec}} - \phi_{\text{true}})], \quad (\text{D1})$$

rather than the direct principal-phase residual used by the fiducial model. The ablation loss is therefore

$$L_{\text{ab}} = L_{\text{time}} + \sigma_{\text{mag}} L_{\text{mag}} + \sigma_{\text{phase}} L_{\text{phase}}^{\text{wrap}}. \quad (\text{D2})$$

Because the same objective is used for both ablation models, the phase convention does not favor either architecture.

We also record the ablation PSNR using the trace-level convention for this comparison. For trace i , containing all three polarization channels, we define

$$\text{PSNR}_{\text{ab},i} = 20 \log_{10} \left[\frac{\max_{c,t} |x_{\text{true},i,c}(t)|}{\sqrt{\text{MSE}_i}} \right], \quad (\text{D3})$$

where MSE_i is averaged over all channels and samples. We then average $\text{PSNR}_{\text{ab},i}$ over the test traces. This convention is specific to the ablation and should not be confused with the per-channel Hilbert-envelope PSNR in Eq. (9).

After selecting the best checkpoint from each run on the validation subset, we evaluate both models once on the same untouched test subset of 41,068 traces using fixed 512-sample windows and no augmentation. The dual-branch model gives $L_{\text{ab}} = 16.314$ and $\langle \text{PSNR}_{\text{ab}} \rangle = 33.738$ dB, compared with $L_{\text{ab}} = 17.574$ and 33.341 dB for the time-only model. Thus, the time-only loss is 7.7% larger, while the Fourier pathways increase the mean test PSNR by 0.40 dB. This paired comparison indicates a modest improvement from the Fourier pathways. With one training realization per model, it does not estimate run-to-run statistical uncertainty.

-
- [1] L. A. Anchordoqui, Ultra-high-energy cosmic rays, *Phys. Rep.* **801**, 1 (2019).
 - [2] R. Alves Batista, The quest for the origins of ultra-high-energy cosmic rays, in *28th European Cosmic Ray Symposium* (2024), [arXiv:2412.17201](https://arxiv.org/abs/2412.17201).
 - [3] J. Álvarez-Muñiz *et al.* (GRAND Collaboration), The Giant Radio Array for Neutrino Detection (GRAND): Science and design, *Sci. China Phys. Mech. Astron.* **63**, 219501 (2020).
 - [4] T. Huege, Radio detection of cosmic ray air showers in the digital era, *Phys. Rep.* **620**, 1 (2016).
 - [5] F. G. Schröder, Radio detection of cosmic-ray air showers and high-energy neutrinos, *Prog. Part. Nucl. Phys.* **93**, 1 (2017).
 - [6] D. Shipilov *et al.*, Signal recognition and background suppression by matched filters and neural networks for Tunka-Rex, *EPJ Web Conf.* **216**, 02003 (2019).
 - [7] C. K. O. Watanabe, P. S. R. Diniz, T. Huege, and J. de Mello Neto, Denoising cosmic rays radio signal using wavelets techniques, *Proc. Sci. ICRC2021* (2021) 260.
 - [8] J. A. Högbom, Aperture synthesis with a non-regular distribution of interferometer baselines, *Astron. Astrophys. Suppl. Ser.* **15**, 417 (1974), <https://ui.adsabs.harvard.edu/abs/1974A%26AS...15..417H/abstract>.
 - [9] A. Horneffer *et al.* (LOPES Collaboration), LOPES: Detecting radio emission from cosmic ray air showers, *Proc. SPIE Int. Soc. Opt. Eng.* **5500**, 129 (2004).
 - [10] S. Martinelli, T. Huege, D. Ravignani, and H. Schoorlemmer, Quantifying energy fluence and its uncertainty for radio emission from particle cascades in the presence of noise, *Astropart. Phys.* **168**, 103091 (2025).
 - [11] C. Gheller and F. Vazza, Convolutional deep denoising autoencoders for radio astronomical images, *Mon. Not. R. Astron. Soc.* **509**, 990–1009 (2022).
 - [12] M. Terris, A. Dabbech, C. Tang, and Y. Wiaux, Image reconstruction algorithms in radio interferometry: From handcrafted to learned regularization denoisers, *Mon. Not. R. Astron. Soc.* **518**, 604–622 (2023).
 - [13] L. Einig *et al.*, Deep learning denoising by dimension reduction: Application to the ORION-B line cubes, *Astron. Astrophys.* **677**, A158 (2023).
 - [14] C. Reissel, S. Soni, M. Saleem, M. Coughlin, P. Harris, and E. Katsavounidis, Toward autonomous denoising of gravitational-wave detector data, *Phys. Rev. D* **113**, 122007 (2026).

- [15] M. Erdmann, F. Schlüter, and R. Smida, Classification and recovery of radio signals from cosmic ray induced air showers with deep learning, *J. Instrum.* **14**, P04005 (2019).
- [16] P. Bezyazeev, D. Shipilov, D. Kostunin, I. Plokhikh, A. Mikhaylenko, P. Turishcheva, S. Golovachev, V. Sotnikov, and E. Sotnikova, Reconstruction of sub-threshold events of cosmic-ray radio detectors using an autoencoder, in *Proceedings of 37th International Cosmic Ray Conference—PoS(ICRC2021)*, ICRC2021 (Sissa Medialab, Trieste, Italy, 2021), p. 223.
- [17] F. G. Schroeder and A. Rehman (IceCube Collaboration), Application of machine learning to identify radio pulses of air showers at the South Pole (ARENA 2024), *Proc. Sci. ARENA2024* (2024) 034. [[arXiv:2502.19931](#)].
- [18] R. Abbasi *et al.* (IceCube Collaboration), Identification and denoising of radio signals from cosmic-ray air showers using convolutional neural networks, *Phys. Rev. D* **113**, 122002 (2026).
- [19] A. Benoit-Lévy, Z. Lai, O. Macias, and A. Ferrière (GRAND Collaboration), Denoising radio pulses from air showers using machine-learning methods, *Proc. Sci. ICRC2025* (2025) 185.
- [20] P. Correa and J. Köhler (GRAND Collaboration), NUTRIG: Development of a novel radio self-trigger for GRAND, *Proc. Sci. ICRC2025* (2025) 229. [[arXiv:2507.04339](#)].
- [21] S. Chiche, GRANDProto300: Status, science case, and prospects, [arXiv:2409.02195](#).
- [22] J. Alvarez-Muñiz, W. R. Carvalho, M. Tüeros, and E. Zas, Coherent Cherenkov radio pulses from hadronic showers up to EeV energies, *Astropart. Phys.* **35**, 287 (2012).
- [23] C. S. C. Sanchez, P. M. Hansen, M. Tüeros, J. Alvarez-Muñiz, and D. G. Melo, Uncertainties in the estimation of air shower observables from Monte Carlo simulation of radio emission, *Astropart. Phys.* **175**, 103183 (2026).
- [24] M. Guelfand (GRAND Collaboration), The Giant Radio Array for Neutrino Detection (GRAND) and its prototype phases, in *58th Rencontres de Moriond on Very High Energy Phenomena in the Universe* (2025), [arXiv:2501.01851](#).
- [25] R. Alves Batista *et al.* (GRAND Collaboration), GRAND-lib: A simulation pipeline for the Giant Radio Array for Neutrino Detection (GRAND), *Comput. Phys. Commun.* **308**, 109461 (2025).
- [26] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, Extracting and composing robust features with denoising autoencoders, in *Proceedings of the 25th International Conference on Machine Learning*, ACM International Conference Proceeding Series Vol. 307 (ACM, New York, 2008), pp. 1096–1103, [10.1145/1390156.1390294](#).
- [27] K. He, X. Zhang, S. Ren, and J. Sun, Identity mappings in deep residual networks, [arXiv:1603.05027](#).
- [28] Z. Lai, O. Macias, and A. Benoit-Lévy, Deep-learning denoising of radio observations for ultra-high-energy cosmic-ray detection, Zenodo (2026), software, `raytune_lib_final` release, [10.5281/zenodo.18233878](#).
- [29] Z. Lai, O. Macias, A. Benoit-Lévy, A. Ferrière, and M. Tüeros, Radio signal denoising with CNN and ray tune, GitHub (2026), software and analysis products, `raytune_lib_final_v1` branch (accessed September 10, 2026), https://github.com/grand-mother/ML_denoising, https://github.com/grand-mother/ML_denoising/tree/raytune_lib_final_v1.
- [30] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, PyTorch: An imperative style, high-performance deep learning library, [arXiv:1912.01703](#).
- [31] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](#).
- [32] L. N. Smith, Cyclical learning rates for training neural networks, [arXiv:1506.01186](#).
- [33] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, Tune: A research platform for distributed model selection and training, [arXiv:1807.05118](#).
- [34] L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, M. Hardt, B. Recht, and A. Talwalkar, A system for massively parallel hyperparameter tuning, [arXiv:1810.05934](#).
- [35] A. Ferrière, S. Prunet, A. Benoit-Lévy, M. Guelfand, K. Kotera, and M. Tüeros, Analytical planar wavefront reconstruction and error estimates for radio detection of extensive air showers, *Nucl. Instrum. Methods Phys. Res., Sect. A* **1072**, 170178 (2025).
- [36] V. Decoene, O. Martineau-Huynh, M. Tüeros, and S. Chiche, A reconstruction procedure for very inclined extensive air showers based on radio signals, *Proc. Sci. ICRC2021* (2021) 211. [[arXiv:2107.03206](#)].
- [37] O. Macias, Z. Mason, M. Ho, A. Ferrière, A. Benoit-Lévy, and M. Tüeros, Simulation-based inference for direction reconstruction of ultra-high-energy cosmic rays with radio arrays, *Phys. Rev. D* **113**, 063018 (2026).
- [38] H. L. Van Trees and K. L. Bell, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*, 2nd ed. (Wiley, Hoboken, NJ, 2013).