

Deep Learning Approaches to Radio Frequency Signal Source Separation

Jared Miller
University of Utah

Tyler Trotter
University of Utah

Hyrum Bailey
University of Utah

Abstract—Signal Source Separation (SSS) in the radio frequency (RF) domain presents a significant challenge due to co-channel interference, diverse modulation schemes, and stochastic noise. This work evaluates RF source separation under a representative linear mixture model of the form $s_m = s_{soi} + \alpha s_{int} + \varepsilon$, where s_{soi} is the signal of interest (SOI), s_{int} is an interfering waveform, and ε represents additive white Gaussian noise (AWGN). Utilizing synthetic quadrature phase shift keying (QPSK) datasets, we characterize performance across single- and multi-channel scenarios under varying noise floors, phase offsets, and interference magnitudes (α). Symbols are encoded using Quadrature Phase Shift Keying (QPSK), corresponding to a four-symbol alphabet. We benchmark several deep learning architectures—including convolutional (IQ_CNN), recurrent (LSTM), and hybrid models—against a FastICA baseline and a modified state-of-the-art audio separation framework (HTDemucs) [1]. Performance is assessed using both waveform-level Mean Squared Error (MSE) and symbol recovery accuracy. Our results reveal a critical decoupling between waveform reconstruction fidelity and symbol decoding success; models achieving minimal MSE do not inherently yield optimal symbol error rates. These findings emphasize the importance of selecting evaluation criteria that align with specific operational objectives, such as bit-error-rate minimization versus high-fidelity signal reconstruction.

I. INTRODUCTION

Signal Source Separation (SSS) is a foundational challenge in digital signal processing with broad implications for wireless communications, spectrum monitoring, and electronic warfare. In its canonical form, SSS involves recovering latent source signals from a composite observation. While the problem is well-characterized in the audio domain—such as in music source separation where individual instruments are extracted from a mixture—the radio frequency (RF) domain introduces unique complexities including high-speed phase variations, diverse modulation schemes, and the requirement for precise temporal synchronization.

The RF environment is increasingly defined by spectral congestion and contested signaling. In these settings, the ability to isolate a signal of interest (SOI) from co-channel interference is critical for maintaining communication reliability and operational awareness. For example, in contested electromagnetic environments, robust source separation can differentiate legitimate transmissions from active jamming or unintentional interference (see Fig. 1).

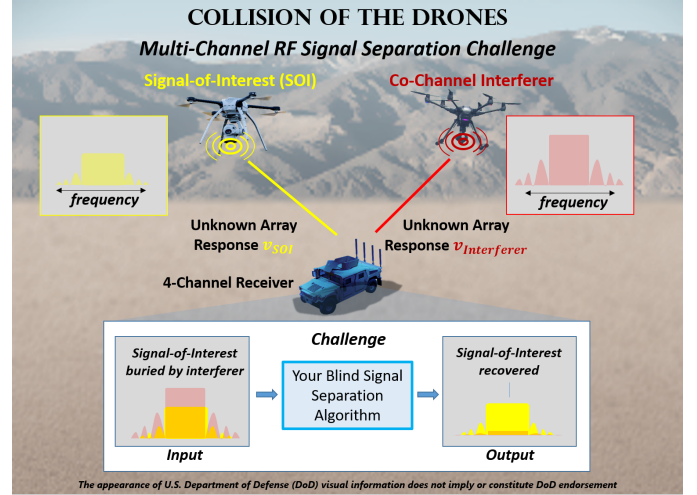


Fig. 1: MIT Radio Frequency Challenge Visualization (Fig. from [2])

This work investigates RF source separation under a representative linear mixture model, a framework common to benchmark efforts such as the MIT Radio Frequency Challenge [2]:

$$s_{\text{mix}} = s_{\text{soi}} + \alpha s_{\text{int}} + \varepsilon \quad (1)$$

where s_{mix} is the observed complex baseband mixture, s_{soi} is the SOI, s_{int} represents the interfering signal, and $\alpha \in \mathbb{R}^+$ is a scaling coefficient defining the Signal-to-Interference Ratio (SIR). The term ε denotes additive white Gaussian noise (AWGN). This formulation encapsulates the primary difficulty of RF SSS: the recovery of structured information from overlapping waveforms in the presence of stochastic noise.

A central objective of this research is to characterize the relationship between waveform reconstruction and information recovery. In digital communication systems, the ultimate goal is often the accurate decoding of transmitted symbols rather than the high-fidelity reconstruction of the carrier waveform. This motivates a critical comparison between *waveform-level MSE* and *symbol recovery accuracy*, identifying scenarios where these two performance metrics diverge.

This work evaluates several deep learning architectures, including convolutional and recurrent neural networks, using synthetically generated Quadrature Phase Shift Keying (QPSK) datasets. We provide a systematic evaluation of these networks

across varying signal-to-noise ratios (SNR) and interference strengths, characterizing their generalization capabilities and identifying specific failure modes in both single-channel and multi-channel separation tasks.

II. BACKGROUND

The research presented herein evaluates source separation within the context of a linear mixture model, defined as:

$$s_{\text{mix}}(t) = s_{\text{soi}}(t) + \alpha s_{\text{int}}(t) + \varepsilon(t) \quad (2)$$

where $s_{\text{soi}}(t)$ and $s_{\text{int}}(t)$ are waveforms generated via Quadrature Phase Shift Keying (QPSK). This section details the signal processing pipeline, from complex baseband representation to symbol recovery.

A. Model and Modulation

Let $\{a_k\}$ denote the transmitted symbol sequence drawn from a four-point QPSK constellation. This symbol sequence is transformed into a continuous-time baseband waveform through pulse shaping:

$$s(t) = \sum_k a_k p(t - kT_s) \quad (3)$$

In this formulation, $p(t)$ represents the pulse-shaping filter and T_s is the symbol period. The corresponding symbol rate is

$$R_s = \frac{1}{T_s} \quad (4)$$

To minimize Inter-Symbol Interference (ISI) while maximizing spectral efficiency, this work utilizes the Root Raised Cosine (RRC) filter for $p(t)$. The foundation of digital RF transmission is the complex baseband signal, represented as:

$$s(t) = I(t) + jQ(t) \quad (5)$$

where $I(t)$ and $Q(t)$ denote the in-phase and quadrature components, respectively. In QPSK modulation, data bits are mapped to a four-point constellation in the complex plane, where each symbol represents a unique phase state. Equivalently, QPSK can be represented as a four-symbol phase alphabet in the complex plane:

$$\mathcal{A} = \{1, j, -1, -j\} \quad (6)$$

where each symbol corresponds to a phase rotation of $\pi/2$ radians. In this representation, each transmitted symbol can be expressed as $a_k \in \mathcal{A}$, linking the discrete bit mapping to its complex exponential phase interpretation. For conceptual clarity, the QPSK alphabet is written here as in Equation 6, which emphasizes the four phase states. In implementation, equivalent normalized constellation points may be used without changing the separation problem, provided the same normalization is applied consistently during data generation, training, and evaluation.

For physical transmission, the baseband signal must be up-converted to a carrier frequency, f_c , to produce the real-valued bandpass signal:

$$x(t) = \text{Re}\{s(t)e^{j2\pi f_c t}\} \quad (7)$$

Expanding this expression yields the standard quadrature modulation form:

$$x(t) = I(t) \cos(2\pi f_c t) - Q(t) \sin(2\pi f_c t) \quad (8)$$

This up-conversion produces the real-valued bandpass signal used for transmission.

B. Receiver Model and Demodulation

The receiver's objective is to reverse the modulation pipeline to reconstruct the original symbol sequence from the captured waveform. This process, termed demodulation, begins with frequency down-conversion. The received bandpass signal is multiplied by a local oscillator at the carrier frequency:

$$y_{\text{base}}(t) = x(t)e^{-j2\pi f_c t} \quad (9)$$

Applying a low-pass filter to $y_{\text{base}}(t)$ removes high-frequency components, recovering the baseband signal $s(t)$.

Following down-conversion, the receiver employs a matched filter—specifically, the time-reversed RRC filter $p(-t)$ —to maximize the signal-to-noise ratio (SNR) at the sampling instances. The output of the matched filter is sampled at intervals of T_s to obtain discrete symbol estimates:

$$\hat{a}_k \approx (\hat{s}(t) * p(-t))|_{t=kT_s} \quad (10)$$

The final stage of the pipeline is the decision logic, or constellation mapping. Each estimate $\hat{a}_k$ is mapped to the nearest ideal constellation point in the QPSK alphabet. This decision stage is critical; it allows the system to tolerate non-deterministic impairments such as thermal noise (ε), the noise generated by agitation of electrons, and residual interference. In the context of source separation, this stage provides the mechanism where bit-error-rate (BER) may remain low even if the separation model achieves sub-optimal waveform reconstruction, as explored in [3].

It is important to note the scope of the present framework. The experiments operate on synthetically generated complex baseband signals and evaluate separation, matched filtering, and symbol decisions within that representation. The framework does not model the full RF front-end pipeline, such as carrier acquisition, frequency offset correction, timing recovery, analog impairments, or over-the-air bandpass transmission and reception. Accordingly, the results should be interpreted as characterizing source separation behavior at the complex baseband level rather than as a complete end-to-end RF communications system.

C. Data Representation for Learning

The learning-based separator operates on real-valued representations of the complex baseband signal. Specifically, the complex waveform in (4) is represented as a two-channel real-valued sequence:

$$\mathbf{x}(t) = \begin{bmatrix} I(t) \\ Q(t) \end{bmatrix} \quad (11)$$

For a sampled signal of length N discrete-time samples, this results in a tensor of shape $\mathbb{R}^{2 \times N}$. In the case of multiple

receive antennas or channels, the input generalizes to $\mathbb{R}^{C \times N}$, where C includes the I/Q components across all receivers.

Similarly, the target outputs (separated sources) are represented as stacked I/Q pairs for each source. For example, separating two sources yields an output tensor of shape $\mathbb{R}^{4 \times T}$ corresponding to:

$$[I_{\text{soi}}, Q_{\text{soi}}, I_{\text{int}}, Q_{\text{int}}]$$

This representation allows standard real-valued neural network architectures to process complex-valued RF signals without requiring specialized complex-valued operations.

III. METHODOLOGY

This section describes the experimental framework developed to evaluate the efficacy of various signal source separation architectures. The system is partitioned into a training-phase pipeline and a execution-phase pipeline, as illustrated in Fig. 26 and Fig. 27, respectively (found in Appendix A).

A. Interface and Configuration Management

The framework provides two primary interfaces for experimental control. The first is a Command-Line Interface (CLI) which enables reproducible experiment specification and supports training, evaluation, and dataset generation through a unified set of runtime flags. These parameters are parsed into a structured configuration object that governs both the data generation pipeline and model execution.

The configuration object controls four main aspects of the experiments: execution mode, model specification, optimization hyperparameters, and synthetic signal-generation settings. These settings include the modulation format, number of transmitted symbols, samples-per-symbol ratio, pulse-shaping parameters, number of receive channels, interference strength α , and noise level. Additional options support checkpoint loading, dataset selection, and user-defined symbol inputs. Together, these parameters determine the mixture-generation process and ensure reproducibility across training and evaluation runs.

In addition to the CLI, a browser-based interface enables interactive evaluation and visualization of trained models. This interface supports qualitative inspection of reconstructed I/Q trajectories and training behavior alongside the quantitative metrics reported in the paper.

B. Synthetic Data Generation

To address the scarcity of labeled, high-fidelity RF datasets for supervised learning, a custom data generation pipeline was developed. The generator produces complex baseband mixtures following the model 2. The pipeline supports a wide range of physical layer configurations, including adjustable symbol counts, samples-per-symbol, and pulse-shaping parameters such as RRC rolloff (β) and filter span. If f_s denotes the sampling frequency and R_s denotes the symbol rate, then the number of samples per symbol is

$$L = \frac{f_s}{R_s} = f_s T_s \quad (12)$$

where T_s is the symbol period. In the primary experiments, this value was typically set to $L = 2$. To ensure numerical stability during training, the generator includes an optional power normalization module. There are two configurations, single-input multiple-output (SIMO) and multiple-input multiple-output (MIMO). For multi-channel (SIMO/MIMO) configurations, the pipeline simulates spatial diversity by applying distinct mixing matrices to the component signals. This allows for rigorous testing under varying noise levels and interference scaling factors (α).

In multi-channel (SIMO/MIMO) configurations, the observed signals are modeled as linear mixtures of the underlying sources through a complex-valued mixing matrix. Let $\mathbf{s}(t) \in \mathbb{C}^K$ denote the vector of K source signals and $\mathbf{x}(t) \in \mathbb{C}^N$ the vector of N received signals. The mixing process is defined as:

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t) \quad (13)$$

where $\mathbf{H} \in \mathbb{C}^{N \times K}$ is the mixing matrix and $\boldsymbol{\varepsilon}(t)$ represents additive noise.

Each entry of $\mathbf{H}$ models the complex channel gain between a source and a receiver, capturing both amplitude scaling and phase shift. In earlier versions of the data pipeline, the entries of $\mathbf{H}$ were sampled with random amplitudes and phases to simulate realistic propagation conditions. While this increases diversity, it also introduces significant variability in the mixing process, leading to a more difficult and less stable learning problem.

In the current configuration, the mixing matrix is constrained to have uniform amplitudes across all source–receiver pairs, while incorporating deterministic phase offsets. Specifically, each channel applies a fixed phase shift to the incoming signals, resulting in a structured mixing matrix of the form:

$$H_{n,k} = e^{j\phi_n} \quad (14)$$

where ϕ_n is a small, channel-dependent phase offset, set to $\phi_n = 5$. This design preserves spatial diversity through phase differences while avoiding the ill-conditioning introduced by random amplitude scaling.

This modification significantly stabilizes training by reducing ambiguity in the mixing process. In particular, when amplitudes and phases are both randomized, the resulting mixing matrix can become poorly conditioned, making inversion or separation more difficult for both learned and classical methods. By contrast, the structured phase-only mixing used here provides a controlled level of diversity while maintaining a more consistent and learnable transformation.

For completeness, the pipeline retains the ability to generate fully randomized mixing matrices with both amplitude and phase variation, enabling evaluation under more realistic but challenging channel conditions when desired.

If a waveform contains M transmitted symbols and uses L samples per symbol, then each discrete-time training example contains

$$N = ML \quad (15)$$

samples per I/Q channel. Throughout the paper, N denotes the total number of discrete-time samples in one example, while M denotes the number of transmitted symbols.

C. Training Configuration

All learned models were trained under a consistent optimization framework to ensure fair comparison across architectures. Unless otherwise noted, the following configuration was used.

Optimizer: Adam optimizer was used for all models.

Learning Rate: The base learning rate was set to 1×10^{-3} , with minor variations explored during hyperparameter tuning for specific models (e.g., IQ_CNN in the multi-channel setting).

Batch Size: Training was performed with a batch size of 16.

Epochs: Models were trained for 20 epochs in the primary experiments, which was sufficient for convergence in the single-channel setting. Longer training was explored for HTDemucs due to slower convergence.

Loss Function: All models were optimized using mean squared error (MSE) between the reconstructed waveform and the ground-truth signal of interest.

Learning Rate Scheduling: For selected experiments, a plateau-based learning rate scheduler was employed, reducing the learning rate when validation loss ceased to improve.

Regularization: Dropout was applied where supported by the architecture, with typical values in the range 0.05–0.1. Weight decay was used in selected configurations (e.g., 10^{-4} scale) to improve generalization.

Gradient Stabilization: Gradient clipping was applied in some experiments (e.g., clipping norm of 2.0) to stabilize training, particularly in recurrent models.

Dataset Size: For the primary single-channel experiments, the dataset consisted of 10,000 training examples and 1,000 validation examples. Each example contained a fixed-length sequence of 400 QPSK symbols.

Data Generation: Unless otherwise specified, training data was pre-generated and fixed for the duration of training to ensure reproducibility. Additional experiments used on-the-fly generation for increased diversity.

Input Representation: All models operate on real-valued I/Q representations of the complex baseband signal, with input tensors of shape $\mathbb{R}^{2 \times T}$ for single-channel and $\mathbb{R}^{2N \times T}$ for N receive channels.

Implementation: Models were implemented in PyTorch and trained on GPU hardware, although no architecture-specific optimizations were required beyond standard deep learning practices.

D. Source Separation Architectures

The framework supports the benchmarking of five distinct separation techniques. These include four deep learning architectures—the *Hybrid Network*, *LSTM Network*, *IQ_CNN*, and the *HTDemucs* framework—and a purely statistical baseline utilizing Fast Independent Component Analysis (FastICA). To address RF source separation, four different network architectures were used, as well as a baseline. The architectures

were Hybrid, LSTM, IQ_CNN, and HTDemucs [4], along with a FastICA baseline. All of the architectures were created (or recreated) to fit the data generation, training, and validation pipelines. Below is a detailed explanation of each method:

1) *FastICA*: In signal processing, independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents. By assumption, at most one subcomponent is distributed according to a Gaussian distribution, and the subcomponents are independent from one another. The goal is to reconstruct observed signals from a number of requisite components by means of a weights matrix. The algorithm finds the independent components by maximizing the independence of the estimated subcomponents [5]. This algorithm classically defines the ‘cocktail party problem,’ where you desire a single person’s voice from a mixture. Formally, you are provided mixed signals, which you assume to be a linear combination of unknown independent source signals which are non-Gaussian:

$$X = WS$$

where X are your observations, W is an unknown mixing matrix, and S are the sources. ICA tries to find an unmixing matrix, W^{-1} such that:

$$S \approx W^{-1}X$$

FastICA iteratively guesses a component direction, measures its non-Gaussianity, adjusts to maximize the independence along that direction, and then removes linear dependence. This serves as a classical linear baseline for our signal source separation pipeline.

2) *Hybrid Separator*: The hybrid separator network combines a linear front-end with a deep temporal network. It begins with input projection. This projection is implemented as a 1×1 convolution that mixes the I and Q input channels to a single feature vector. This enables channel feature fusion. After the projection layer is done, the model uses a series of convolutional blocks. Each block has a kernel size of 1×9 . This allows the model to learn structure over time which is critical for the multi-channel separation problem. The last portion of the network is another 1×1 convolution that re-projects the learned representation back to our original desired output representation which are the estimated separated signals. This model acted as our baseline for the multi-channel separation problem.

3) *LSTM Separator*: A Long Short-Term Memory (LSTM) network is a specialized Recurrent Neural Network (RNN) designed to recognize patterns in data sequences, such as text, speech, and time-series data. An LSTM works by maintaining a memory state c_t , using gates to control information flow, enabling stable long-term gradient propagation.

In this network for every example in the batch both the I and Q values received at each antenna are used. This allows the model to see the full instantaneous observation across all antennas. We then do a simple transpose to make time the sequence dimension. Without this transpose the network

wouldn't be able to utilize its memory effectively. We then do an input projection and pass it into an LSTM.

4) *IQ_CNN Separator*: The IQ_CNN separator is a U-Net style temporal convolutional model for RF source separation. The model operates on the received complex baseband signal after demodulation, where each receive channel is represented by its in-phase and quadrature components. Thus, for n_{rx} receive channels, the network input is arranged as $2n_{rx}$ real-valued channels, corresponding to the I and Q values for each receiver. This allows the model to process the same information contained in the complex baseband waveform while remaining compatible with standard real-valued neural network layers. Although the received RF waveform is real-valued, it is digitally downconverted to complex baseband prior to processing. This is achieved by mixing the signal with cosine and sine carriers at the center frequency and applying low-pass filtering, producing I and Q components.

Each convolutional block consists of two sequential 1D convolutions with group normalization and GELU activation. The first layer is a 1×1 convolution, which performs a learned linear projection across the channel dimension at each time step. Rather than simply scaling individual channels, this operation computes weighted combinations of all input I/Q channels, enabling the network to mix information across antennas and between in-phase and quadrature components. This can be interpreted as a learned spatial filtering or beamforming operation that maps the raw observations into a more useful feature space for subsequent processing. This operation is equivalent to applying a time-independent fully connected layer to each temporal sample.

After this projection, the network follows an encoder-decoder structure inspired by U-Net [6]. The encoder consists of three downsampling convolutional blocks implemented using stride-2 convolutions, which reduce time resolution while increasing feature depth. This allows the model to capture broader temporal context than the linear or shallow convolutional models. At the bottleneck, the model applies dilated convolutions with dilation rates of 1, 2, and 4, which increase the receptive field without additional loss of resolution [7]. The decoder then upsamples the learned representation back to the original temporal resolution using transposed convolutions and skip connections implemented via channel-wise concatenation, allowing fine waveform detail from earlier layers to be combined with the larger-scale contextual representation.

Group normalization [8] is used throughout the network to stabilize training, especially for moderate or small batch sizes, and GELU activations [9] provide a smooth nonlinear transformation. Optional dropout is included for regularization. The final layer maps the learned features back into two separate complex sources represented in I/Q form. In practice, this model serves as a stronger nonlinear separator than the linear baseline and is intended to capture both the temporal structure of the RF waveform and the multichannel relationships across receive observations.

5) *HTDemucs Separator*: The Hybrid Transformer Deep Extractor for Music Sources (HTDemucs) is a state-of-the-

art open-source audio source separation model developed by Meta Research (Meta AI). The neural network architecture baseline of Demucs is that of a Wave-U-Net and was inspired by the SING synthesis model, including a convolutional encoder stream, a central bidirectional long short-term memory (bi-LSTM) bottleneck, followed by a convolutional decoder with transposed convolutions [1].

The first component consists of a depth of L that progressively downsamples the waveform, increasing feature channels, with a large kernel and stride, followed by a gated linear unit (GLU) [10]. It has shown to significantly outperform previous waveform-based models, for audio, such as D3Net and Wave-U-Net by at least 1.6 dB signal-to-distortion ratio (SDR), see Section III-E below, on the MusDB benchmark [11]. The hybrid transformer is a dual-branch structure with cross-attention between time-domain and frequency-domain representations of the waveform. It takes two inputs: the waveform directly (time domain) and a Fourier transform based representation (frequency domain) [1] (see Music Source Separation in Appendix C).

For our implementation, we compute a real-valued fast Fourier transform (RFFT) across the entire waveform to obtain a frequency-domain representation. This representation is provided to the frequency branch of the model, while the original waveform is processed by the time-domain branch. The model operates on fixed-length discrete-time I/Q sequences of length N , where N denotes the number of discrete-time samples per training example. Each symbol is oversampled into multiple time-domain samples via pulse shaping. Each training example consists of a fixed number of QPSK symbols, which are pulse-shaped (RRC) into a longer discrete-time I/Q sequence sampled at multiple samples per symbol.

We re-implemented the HTDemucs architecture from scratch and modified the internal dimensions to fit our RF wave separation goals. The input consists of a batch size dimension, B , which is the number of samples processed in parallel, two channels corresponding to the I and Q components, with T time samples. Therefore, the input is $x \in \mathbb{R}^{B \times 2 \times N}$, representing a complex baseband signal using real-valued channels. This network does not use transfer learning from the pre-trained Meta HTDemucs model, and instead the network is trained from scratch on synthetic RF data.

E. Performance Evaluation Metrics

Performance is quantified through two complementary metrics designed to capture both signal fidelity and information integrity:

- **Mean Squared Error (MSE)**: This waveform-level metric is computed on the real-valued I/Q representation of the complex baseband signal. If the ground-truth signal is represented by samples (I_n, Q_n) and the reconstruction by $(\hat{I}_n, \hat{Q}_n)$ for $n = 1, \dots, N$, then

$$\text{MSE} = \frac{1}{2N} \sum_{n=1}^N \left[(I_n - \hat{I}_n)^2 + (Q_n - \hat{Q}_n)^2 \right] \quad (16)$$

This provides a measure of how accurately the full complex baseband waveform is reconstructed. We use IQ-MSE rather than an error metric on the quadrature component alone because the separation objective is defined on the full complex baseband waveform. Although Q-only error can be useful as a diagnostic when phase-sensitive behavior is of interest, restricting the loss to only one component would ignore reconstruction error in the in-phase channel and would no longer measure fidelity of the complete recovered signal.

- **Symbol Recovery Accuracy:** This application-specific metric evaluates the post-demodulation integrity of the data. It is defined as the ratio of correctly decoded symbols to the total number of transmitted symbols:

$$\text{Acc}_{\text{sym}} = \frac{1}{N} \sum_{k=1}^N \mathbb{I}(\hat{a}_k = a_k) \quad (17)$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Several additional metrics were considered throughout experimentation:

Signal-to-Distortion Ratio (SDR): In addition to MSE, we evaluate reconstruction quality using the Signal-to-Distortion Ratio (SDR), which measures the ratio of the desired signal power to the residual distortion in the reconstructed waveform. In this work, distortion is defined as the difference between the ground-truth signal of interest and its estimate.

Formally, SDR is computed as:

$$\text{SDR} = 10 \log_{10} \left(\frac{\|s_{\text{soi}}(t)\|^2}{\|s_{\text{soi}}(t) - \hat{s}_{\text{soi}}(t)\|^2} \right) \quad (18)$$

where $s_{\text{soi}}(t)$ is the ground-truth signal and $\hat{s}_{\text{soi}}(t)$ is the reconstructed signal. Higher SDR values indicate lower distortion and better reconstruction fidelity.

This formulation corresponds to a reconstruction-based SDR and is closely related to inverse normalized mean squared error. Unlike classical definitions of SDR that explicitly decompose interference, noise, and artifacts, this formulation provides a compact measure of overall reconstruction quality. While SNR governs the impact of stochastic noise, the interference scaling parameter α controls the relative strength of structured interference. Together, these parameters define the overall difficulty of the separation task.

Signal-to-Noise Ratio (SNR): To characterize channel conditions, we also consider the Signal-to-Noise Ratio (SNR), which quantifies the relative strength of the signal of interest to the additive noise component. In the context of the mixture model, SNR is defined as:

$$\text{SNR} = 10 \log_{10} \left(\frac{\|s_{\text{soi}}(t)\|^2}{\|\varepsilon(t)\|^2} \right) \quad (19)$$

where $\varepsilon(t)$ represents additive white Gaussian noise (AWGN). Higher SNR values correspond to cleaner observations with less noise corruption.

In the experimental pipeline, SNR is controlled either directly via a specified SNR level in decibels or indirectly through the

variance of the additive noise. This allows systematic evaluation of model robustness under varying noise conditions.

A central premise of this work is that while a network may exhibit non-trivial MSE (particularly on validation sets), it may still maintain the structural properties necessary for high-accuracy symbol recovery, which is often the primary objective in operational communication systems.

IV. TRAINING

All learned models were trained using the Adam optimizer on fixed synthetic datasets generated by the RF data pipeline. For the finalized single-channel comparison, the dataset consisted of 10,000 training examples and 1,000 validation examples. All models (Hybrid, LSTM, IQ_CNN, and HTDemucs) were optimized using a plain mean squared error (MSE) objective to ensure a consistent comparison under the same training criterion.

Under this setup, Hybrid, LSTM, and IQ_CNN converged rapidly and reached stable performance within 20-epochs. In contrast, HTDemucs exhibited significantly slower convergence and required substantially longer training to approach similar performance levels. This difference in training dynamics is shown in Fig. 2, where HTDemucs continues to improve gradually while the other models stabilize early.

The multichannel setting introduced additional complexity due to the use of a channel mixing matrix that applies different amplitude scalings and phase shifts to each received signal. This results in each observation being a distinct linear combination of the underlying sources, increasing the difficulty of the separation problem relative to the single-channel case. In the implementation used here, the mixing matrix initially assigned random amplitudes and phases, which further increased variability in the training data.

To address this increased difficulty, a separate cross-validation effort was conducted for the IQ_CNN model. An initial broad hyperparameter search was performed in the four-channel setting, varying base channel width, dropout, learning rate, weight decay, input noise level, receiver dropout probability, gradient clipping threshold, scheduler type, batch size, and normalization strategy. The strongest configuration from this search used `base_channels = 48`, `dropout = 0.08`, `learning rate = 1.6887 \times 10^{-3}`, `weight decay = 1.315 \times 10^{-4}`, `input-noise standard deviation = 0.005`, `receiver dropout probability = 0.1`, `gradient clipping = 2.0`, and a plateau-based learning rate scheduler. A subsequent localized search around this region further improved validation PIT-MSE, producing the final IQ_CNN configuration used in the multichannel experiments.

For the experimental sweeps reported in this work, the training pipeline was held fixed while varying only the data conditions. In the interference-strength (α) sweep, the optimizer settings, dataset size, and training budget were kept constant while α was varied. In the noise sweep, α was fixed at 0.8 and the variance of additive Gaussian noise was increased. HTDemucs was excluded from these sweeps due to its substantially slower training dynamics and tendency

toward gradual overfitting under extended training, which made direct comparison under the same fixed training budget less meaningful. This behavior is illustrated in Fig. 2.

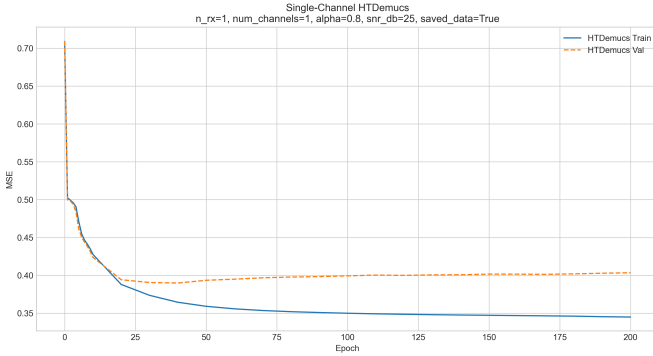


Fig. 2: Single-channel HTDemucs training and validation curves. Compared with Hybrid, LSTM, and IQ_CNN, HTDemucs improved much more slowly and did not catch up within the same short budget.

Please refer to Appendix B for training and validation plots containing each network for each of 1, 2 and 4 channel scenarios.

V. EXPERIMENTS

This section presents experimental results for the single-channel and two-channel separation problems. In this work, the number of channels corresponds to the number of receive antennas. Increasing the number of channels introduces additional complexity because each antenna observes a different linear mixture of the underlying sources, with distinct amplitude and phase offsets.

A. Single-Channel Separation Problem

The single-channel experiments used a fixed configuration with one receive channel ($n_{rx} = 1$, $\text{num_channels}=1$), a saved dataset (rather than on-the-fly generation), no additive noise in the baseline comparison, and an interference strength of $\alpha = 0.8$. Each sequence consisted of a fixed number of 400 symbols (as defined in the data generation pipeline), ensuring consistency across all models.

Under this configuration, Hybrid, LSTM, and IQ_CNN achieved near-perfect symbol recovery and very low waveform error, while HTDemucs improved more slowly and remained significantly behind within the same training budget. The quantitative comparison is shown in Table I, and a representative waveform comparison is shown in Fig. 3. These results establish a strong baseline performance level for the learned models in the single-channel setting.

The interference-strength (α) sweep provides further insight into model behavior. Starting from the baseline $\alpha = 0.8$, increasing α toward the equal-strength case ($\alpha = 1.0$) produces a sharp degradation in both waveform reconstruction and symbol accuracy. As shown in Table II, all three learned models exhibit low error at $\alpha = 0.80$ and $\alpha = 0.95$, but

TABLE I: Single-channel reference comparison at the baseline configuration.

Model	Wave MSE	SDR (dB)	Symbol Acc
LSTM	0.000008	47.70	1.000
IQ_CNN	0.000029	42.49	1.000
Hybrid	0.000081	37.90	1.000
HTDemucs	0.428018	0.68	0.397

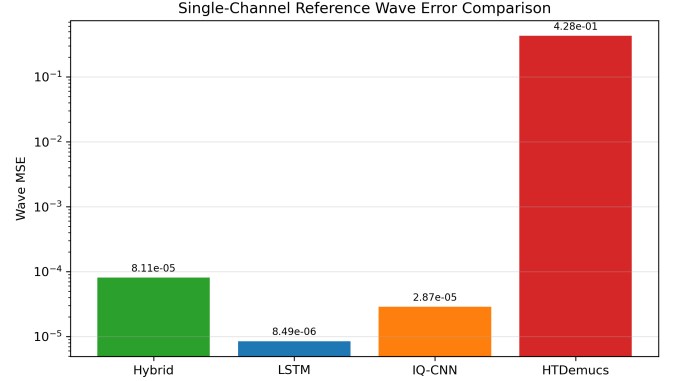


Fig. 3: Representative waveform reconstruction for the single-channel baseline configuration.

performance collapses at $\alpha = 1.0$, where validation MSE rises to approximately 0.25 and symbol accuracy drops to roughly 57%.

1) *Symmetry and Ill-Posedness: $\alpha = 1$:* The degradation observed near $\alpha = 1$ can be understood as a consequence of a fundamental symmetry in the mixture model. Recall the single-channel formulation:

$$s_{\text{mix}} = s_{\text{soi}} + \alpha s_{\text{int}} \quad (20)$$

At the critical point $\alpha = 1$, this reduces to:

$$s_{\text{mix}} = s_{\text{soi}} + s_{\text{int}} \quad (21)$$

Under the data generation procedure, s_{soi} and s_{int} are drawn from identical distributions (QPSK-modulated waveforms with identical pulse shaping, symbol rates, and normalization). Consequently, the joint distribution is exchangeable:

$$(s_{\text{soi}}, s_{\text{int}}) \stackrel{d}{=} (s_{\text{int}}, s_{\text{soi}}) \quad (22)$$

This implies that the observation s_{mix} is invariant under permutation of the latent sources. In particular, for any observed mixture, both decompositions

$$s_{\text{mix}} = s_1 + s_2 \quad \text{and} \quad s_{\text{mix}} = s_2 + s_1 \quad (23)$$

are equally valid and indistinguishable from the observation alone.

a) *Ill-posedness:* Let the inverse problem be defined as recovering s_{soi} from s_{mix} . A problem is well-posed in the sense of Hadamard only if the solution is (i) existent, (ii) unique, and (iii) stable with respect to perturbations in the input.

At $\alpha = 1$, uniqueness is violated. Formally, the conditional distribution satisfies:

$$p(s_{\text{soi}} | s_{\text{mix}}) = p(s_{\text{int}} | s_{\text{mix}}) \quad (24)$$

due to symmetry of the generative model. Therefore, the posterior over sources is permutation-invariant, and no measurable function of s_{mix} can distinguish the signal of interest from the interference.

As a result, the inverse mapping $s_{\text{mix}} \mapsto s_{\text{soi}}$ is not identifiable at $\alpha = 1$ within the present experimental formulation, and the separation problem becomes locally ill-posed with respect to source identity.

b) Consequence for MSE optimization: Under a mean-squared error objective, the optimal estimator is given by the conditional expectation:

$$\hat{s}_{\text{soi}} = \mathbb{E}[s_{\text{soi}} | s_{\text{mix}}] \quad (25)$$

However, due to exchangeability, this estimator collapses to a symmetric solution that does not preserve source identity. For normalized QPSK signals, this results in an effective averaging behavior between indistinguishable components, yielding an irreducible reconstruction error on the order of:

$$\mathbb{E}[\|s_{\text{soi}} - \hat{s}_{\text{soi}}\|^2] \approx 0.25 \quad (26)$$

consistent with the empirical error plateau observed at $\alpha = 1$.

c) Interpretation: This degeneracy explains the non-monotonic behavior observed in the α sweep. The difficulty at $\alpha = 1$ does not arise from increased interference power, but from a symmetry-induced loss of discernibility in the source decomposition. For $\alpha \neq 1$, this symmetry is broken, restoring a well-defined separation objective and enabling successful recovery by the learned models.

Empirically, performance improves again as α moves away from this symmetric point. By $\alpha = 1.05$, all three models recover to near-perfect symbol accuracy, confirming that the degradation is concentrated near the equal-strength regime rather than increasing monotonically with interference power.

TABLE II: Selected points from the single-channel alpha sweep, including both waveform error and symbol accuracy.

α	Hybrid		LSTM		IQ_CNN	
	MSE	Sym Acc	MSE	Sym Acc	MSE	Sym Acc
0.80	0.000165	1.0000	0.000019	1.0000	0.000040	1.0000
0.98	0.000797	0.9998	0.000696	0.9996	0.125624	0.7403
1.00	0.249567	0.5738	0.249504	0.5733	0.249526	0.5730
1.02	0.089617	0.8713	0.000605	0.9996	0.022320	0.9993
1.05	0.000143	1.0000	0.000085	1.0000	0.000154	1.0000

Figures 4 and 5 summarizes this behavior across the full sweep. The peak in error near $\alpha \approx 1.0$ confirms that the most challenging case is not simply high interference, but the symmetric mixture itself.

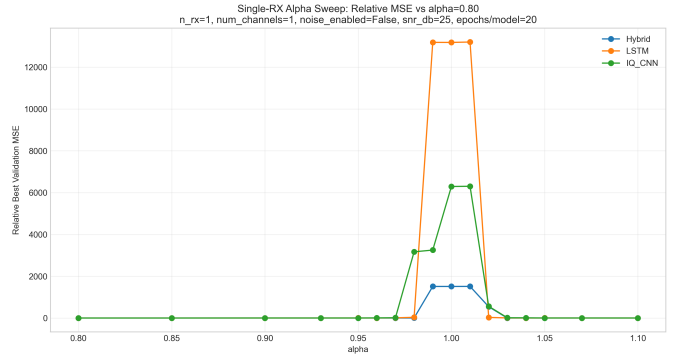


Fig. 4: Relative wave MSE vs. α

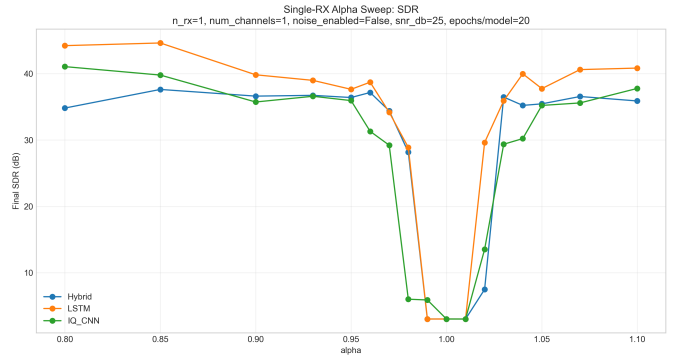


Fig. 5: SDR vs. α

Fig. 6 provides a qualitative comparison of waveform reconstruction at a well-performing point ($\alpha = 0.80$) and Fig. 7 demonstrates the failure point ($\alpha = 1.00$). The degradation at $\alpha = 1.00$ is visually apparent and aligns with the quantitative metrics.

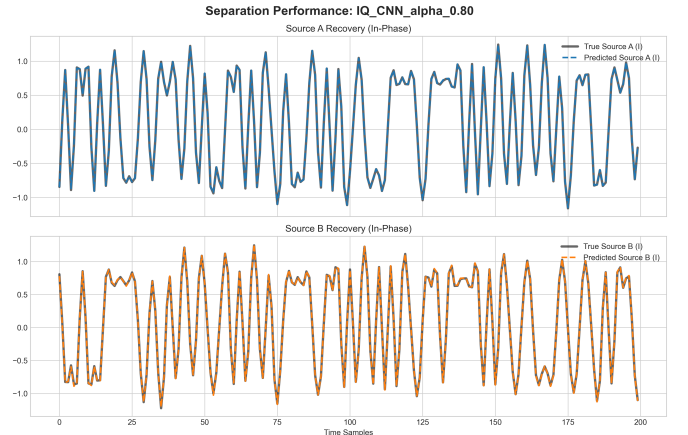


Fig. 6: $\alpha = 0.80$

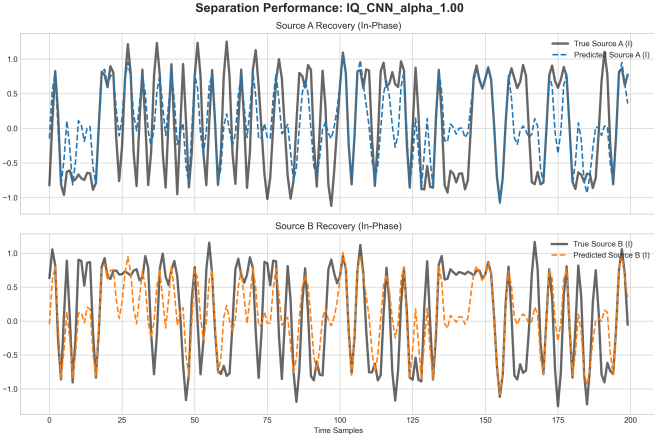


Fig. 7: $\alpha = 1.00$

To further characterize the single-channel setting, we evaluated the separator over a two-dimensional grid of interference strengths and noise powers using the mixture model in (2). Because only a single observation is available, this setting remains fundamentally underdetermined, and successful recovery depends on learned statistical structure rather than on a uniquely invertible mixing process.

The joint sweep over α and σ^2 shows that the single-channel model operates in distinct performance settings. At low noise levels, symbol recovery remains high across a broad range of interference strengths. As σ^2 increases, however, performance degrades rapidly and then stabilizes in a lower-accuracy setting. This behavior is summarized in Table III and visualized in Fig. 8. The surface complements the earlier one-dimensional α -sweep: the earlier sweep isolated the symmetry-induced failure near $\alpha = 1$, while the two-dimensional sweep shows that noise introduces a second axis of failure and that the interaction between interference strength and noise produces a clear transition from near-perfect recovery to partial collapse.

TABLE III: Representative single-channel SOI symbol correctness (%) over the (α, σ^2) grid.

$\alpha \backslash \sigma^2$	0	1	5	10	15	20
0.0	100.0	89.6	58.8	48.2	44.6	41.3
0.5	100.0	82.6	51.8	48.3	40.4	40.8
1.0	83.8	58.2	43.6	41.7	36.6	36.2
1.5	100.0	65.4	37.2	37.0	33.8	32.7
2.0	100.0	76.5	41.2	34.5	31.4	35.5

Table III makes three features clear. First, in the absence of noise, recovery is nearly perfect for almost all interference levels, with the main exception occurring near the equal-strength setting already discussed earlier. Second, even modest noise leads to substantial degradation. Third, as noise increases further, the model does not collapse to random guessing; instead, it stabilizes in a lower-performance setting, indicating that some signal structure is still being exploited.

The corresponding error-count sweep further reinforces this interpretation. While symbol correctness decreases smoothly in the low- to moderate-noise scenario, the number of incorrect

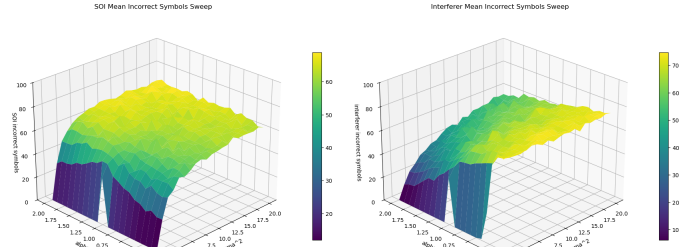


Fig. 8: Mean single-channel SOI symbol correctness as a function of interference strength α and noise power σ^2 .

symbols can become highly variable in the more difficult regions of the grid. This indicates that failure in the single-channel case is not simply a uniform loss of performance, but rather a transition into an unstable setting in which specific noise realizations can strongly affect the decoded sequence. Thus, the single-channel separator should be understood as operating near a statistical decision boundary rather than as performing deterministic inversion.

We also tested sensitivity to phase shifts in the interference signal. The results are summarized in Table IV. In the noiseless setting, performance is essentially perfect across all tested phase shifts. Even when small variation appears at a few angles, the accuracy remains above 99.9% for both the signal of interest and the interferer.

TABLE IV: Single-channel source separation performance versus phase shift of the interference signal. Means are reported as symbol correctness (%), with variance computed across repeated trials.

Phase (deg)	SOI Mean	SOI Var	Int Mean	Int Var
0	100.00	0.00	100.00	0.00
30	100.00	0.00	100.00	0.00
60	100.00	0.00	100.00	0.00
90	100.00	0.00	100.00	0.00
120	100.00	0.00	100.00	0.00
150	100.00	0.00	100.00	0.00
180	100.00	0.00	100.00	0.00
210	100.00	0.00	100.00	0.00
240	99.90	0.09	99.90	0.09
270	100.00	0.00	99.90	0.09
300	100.00	0.00	99.90	0.09
330	100.00	0.00	100.00	0.00
360	99.90	0.09	99.90	0.09

These phase-shift results indicate that the learned separator is largely phase-invariant in the single-channel setting. That is, the model is not depending on a particular relative phase alignment between the signal of interest and the interference in order to recover symbols correctly. This is an important contrast with the much stronger dependence on noise power seen in Fig. 8. In this setting, changes in symbol correctness are driven far more by stochastic corruption than by deterministic phase offsets of the interferer.

This trend is also visible in Fig. 9, where the mean performance remains essentially flat and the shaded variance

band is very small. The figure therefore supports the conclusion that interference phase shift is not a major source of failure for the single-channel separator.

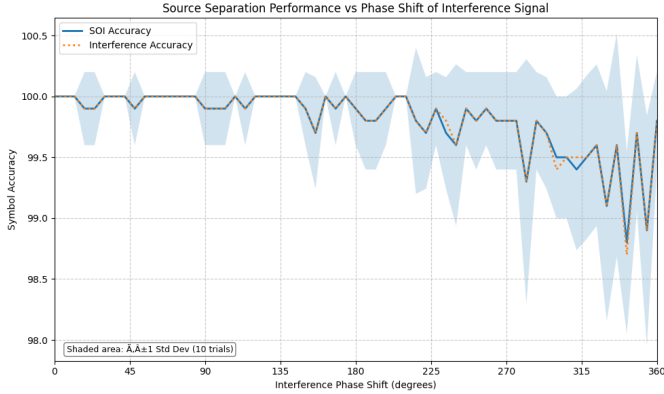


Fig. 9: Single-channel source separation performance versus phase shift of the interference signal. The line shows mean symbol correctness, and the shaded region shows one standard deviation across repeated trials.

We next examined the effect of sampling frequency. Because the model is trained at a fixed samples-per-symbol configuration, changing the sampling frequency at inference time introduces a distribution shift even when the underlying modulation remains unchanged. Representative results are shown in Table V, and the overall trend is visualized in Fig. 10.

TABLE V: Single-channel source separation performance versus sampling frequency. Means are reported as symbol correctness (%), with variance computed across repeated trials.

f_s	SOI Mean	SOI Var	Int Mean	Int Var
1	52.10	17.29	43.70	36.21
2	100.00	0.00	100.00	0.00
10	92.30	12.41	77.20	21.76
20	90.50	5.45	76.30	16.01
40	90.80	2.56	76.30	11.81
60	89.90	3.29	74.00	18.00
100	91.10	2.89	75.40	9.64

The sampling-frequency experiment shows that performance is maximized when the inference-time sampling frequency matches the training configuration ($f_s = 2$). Reducing the number of samples per symbol below that level severely degrades performance, which is expected because the network no longer receives enough temporal resolution to preserve the learned waveform structure. Increasing the sampling frequency above the training value also lowers performance, although in a less catastrophic way. After the initial drop from $f_s = 2$, the curves plateau rather than improving. This indicates that the network has learned features tied to the training distribution rather than a scale-invariant representation of the signal. In this setting, higher sampling resolution is therefore not inherently better; matching the training distribution matters more.

Finally, we evaluated performance as a function of the timing parameters T_1 and T_2 , which control aspects of the

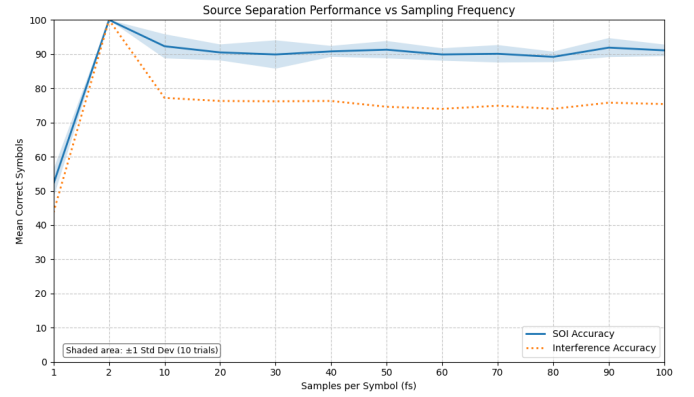


Fig. 10: Single-channel source separation performance versus sampling frequency. The line shows mean symbol correctness, and the shaded region shows one standard deviation across repeated trials.

temporal structure of the generated signals prior to mixing. Unlike the phase-shift experiment, which shows near invariance, the T_1, T_2 sweep reveals that the separator is sensitive to temporal structure in the observation window. This is expected: phase offsets change the orientation of a waveform in the complex plane, whereas changes in T_1 and T_2 alter the temporal arrangement of samples and therefore directly affect how recognizable the waveform is to the learned model.

Fig. 11 shows that the dependence on T_1 and T_2 is more gradual than the sharp transitions induced by noise, but still non-trivial. This provides another example of distribution mismatch in the single-channel problem: the model is comparatively robust to phase perturbations, but more sensitive to changes in temporal structure.

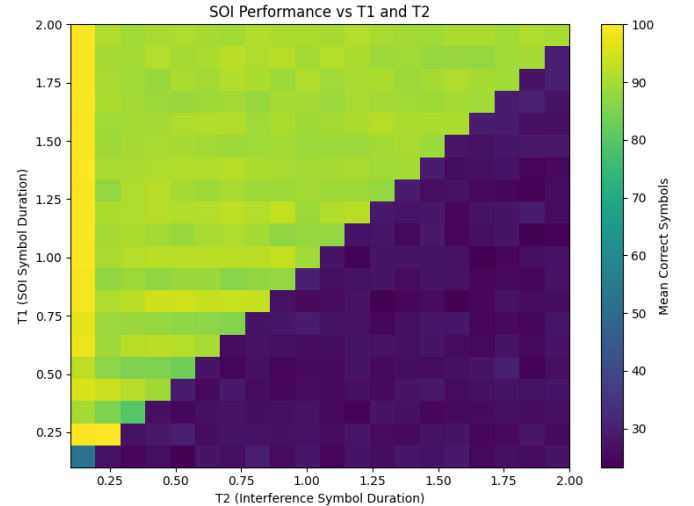


Fig. 11: Single-channel SOI performance versus timing parameters T_1 and T_2

Overall, these additional experiments clarify several important characteristics of the single-channel setting. First,



Fig. 12: Validation MSE over training epochs for the two-channel setting.

performance is highly sensitive to noise, exhibiting a rapid transition from near-perfect recovery to partial failure as σ^2 increases. Second, the model retains meaningful structure even in degraded settings, outperforming random guessing even when recovery is no longer reliable. Third, the separator is comparatively robust to phase perturbations but sensitive to distribution mismatch in temporal sampling and timing structure.

Taken together, these results emphasize that single-channel source separation relies heavily on learned statistical regularities and is inherently limited by both noise and mismatch between training and inference conditions. These findings also reinforce a broader theme of the paper: signal-level reconstruction quality, symbol-level correctness, and robustness to distribution shift must all be examined jointly in order to fully characterize separator performance.

B. Dual-Channel Separation Problem

The two-channel experiments extend the problem to multiple receive antennas, where each observation is a different linear mixture of the underlying sources. The evaluation includes both learned models (Hybrid, LSTM, IQ_CNN, HTDemucs) and the FastICA baseline, which does not require training.

Table VI summarizes the results. A consistent gap between training and validation symbol accuracy is observed for several learned models (e.g., LSTM: 0.9436 training vs. 0.3342 validation), indicating substantial overfitting in the two-channel setting. This trend is also reflected in the validation MSE curves shown in Fig. 12.

TABLE VI: Model performance for the two-channel separation problem under sophisticated mixing matrix scenario.

Model	Final MSE	Train Sym Acc	Val Sym Acc
FastICA	1.48406	0.3304	0.3353
Hybrid	0.50001	0.2598	0.2592
LSTM	0.79414	0.9436	0.3342
IQ_CNN	0.71724	0.6549	0.3393
HTDemucs	0.56651	0.4316	0.2556

These results show that waveform reconstruction error and symbol recovery are not strictly correlated. For example, the Hybrid model achieves the lowest MSE but the lowest symbol accuracy, while FastICA achieves comparable validation symbol accuracy without competitive waveform reconstruction. This highlights the need to evaluate both metrics when assessing separation performance.

Among the learned models, LSTM achieves the highest training accuracy but exhibits significant overfitting, while IQ_CNN provides the strongest validation symbol accuracy with a smaller gap between training and validation performance. This suggests that IQ_CNN generalizes more effectively under the evaluated conditions.

a) Implications for Multi-Channel Design: The ill-posedness observed at $\alpha = 1$ highlights a broader issue in the formulation: successful separation depends critically on the presence of identifiable structure in the mixing process. In the absence of such structure, the inverse problem collapses into a permutation-symmetric decomposition with no unique solution.

This observation directly motivated a simplification of the multi-channel mixing model. In a fully general formulation, the mixing matrix $\mathbf{H} \in \mathbb{C}^{N \times K}$ may include both amplitude and phase variability across channels. However, unconstrained random amplitude scaling can introduce excessive conditioning variability, which obscures the role of spatial diversity in separation performance.

To isolate the effect of spatial structure while preserving identifiability, we constrain the mixing matrix to a phase-only form:

$$H_{n,k} = e^{j\phi_n}, \quad (27)$$

where ϕ_n is a fixed channel-dependent phase offset shared across sources. This formulation removes amplitude-induced conditioning effects while retaining deterministic inter-channel phase diversity.

Importantly, this design preserves the minimal structure required for multi-channel separation: distinct observations remain non-collinear in the complex domain, ensuring that the system is identifiable up to permutation and global phase ambiguity. At the same time, it avoids artificially increasing task difficulty through random amplitude scaling that does not correspond to additional separability information.

Our results illustrate the severity of this multi-channel ambiguity when amplitude is unconstrained. As visualized in the data pipeline mixture (Fig. 13), the overlapping waveforms lack sufficient spatial separability for the networks to uniquely assign sources without overfitting to the training distribution.

The training and validation dynamics further highlight this struggle. As detailed in Appendix B (Fig. 30 and Fig. 31), the learned models—particularly the LSTM—rapidly minimize training error but fail to generalize, leading to the massive divergence between training and validation accuracy seen in Table VI. This confirms that without the phase-only constraint, the deep learning architectures memorize the random

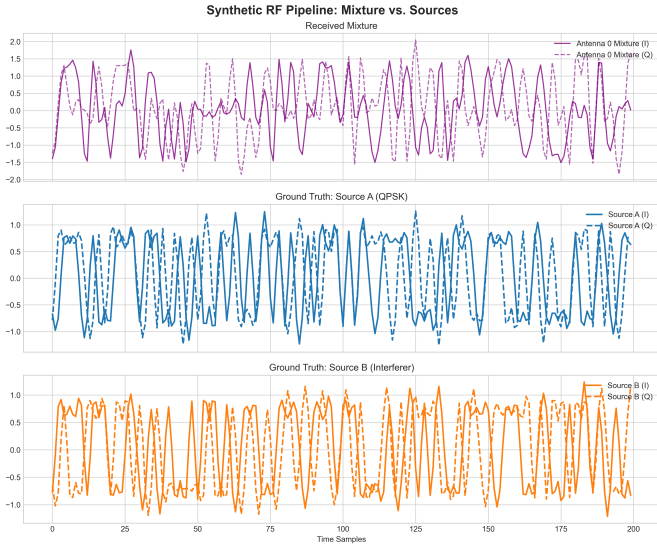


Fig. 13: Data pipeline visualization for the dual-channel configuration, highlighting the complex mixture of unconstrained amplitude and phase.

mixing parameters rather than learning a generalized separation function.

To further characterize the dual-channel setting, under the sophisticated mixing matrix circumstance, we examined performance under several additional perturbations: a joint sweep over interference strength α and noise variance σ^2 , a phase-shift sweep on the interference signal, a sweep over the samples-per-symbol parameter f_s , and a timing-mismatch study over T_1 and T_2 . These experiments were designed to determine whether the poor generalization observed in the unconstrained dual-channel setting is driven primarily by additive noise, by representation mismatch, or by a more structural discernibility issue in the mixing process.

The joint (α, σ^2) sweep is shown in Fig. 14, where performance is reported in terms of mean incorrect symbols for both the SOI and the interferer. Several trends are immediately apparent. First, the dual-channel separator does not exhibit the same near-perfect plateau observed in the single-channel case. Even at low noise, symbol errors remain substantial over a broad region of the sweep. Second, performance degrades as σ^2 increases, but the dependence on α is not monotone. Instead, there are broad regions of elevated error, indicating that the model is sensitive not only to noise strength but also to how the two sources interact through the unconstrained mixing process. Third, the SOI and interferer surfaces are not identical, which is consistent with the fact that the two recovered sources are not equally easy to assign once the model begins to overfit random channel realizations. In other words, the dominant failure mode in the dual-channel setting is not simply noise sensitivity, but instability in source assignment under varying mixtures.

Fig. 14 therefore reinforces the main conclusion of this section: the unconstrained dual-channel problem is difficult

because the model does not learn a robust inverse mapping across channel realizations. Instead, it appears to exploit training-specific structure that does not transfer cleanly across the (α, σ^2) plane.

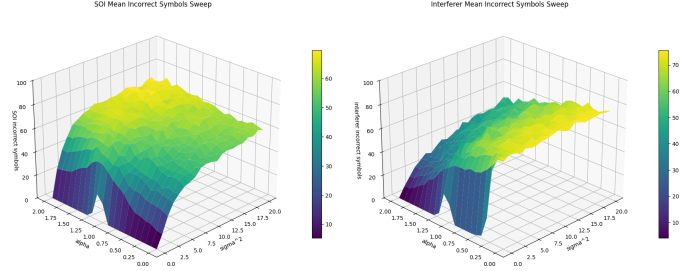


Fig. 14: Mean incorrect symbols in the dual-channel setting as a function of interference strength α and noise variance σ^2 , shown separately for the SOI and the interferer.

We also evaluated sensitivity to the phase of the interference signal. Fig. 15 shows that symbol accuracy remains extremely high across the full phase-shift range, with only very small fluctuations. This indicates that, under otherwise matched conditions, the learned separator is largely insensitive to interference phase alone. The shaded region is narrow, and the mean curves remain essentially flat, so the small differences that do appear are more plausibly attributed to random trial-to-trial variation than to any systematic phase-dependent failure mode. This is an important contrast with the broader dual-channel failure patterns discussed above: phase rotation by itself is not the main reason the unconstrained dual-channel model generalizes poorly.

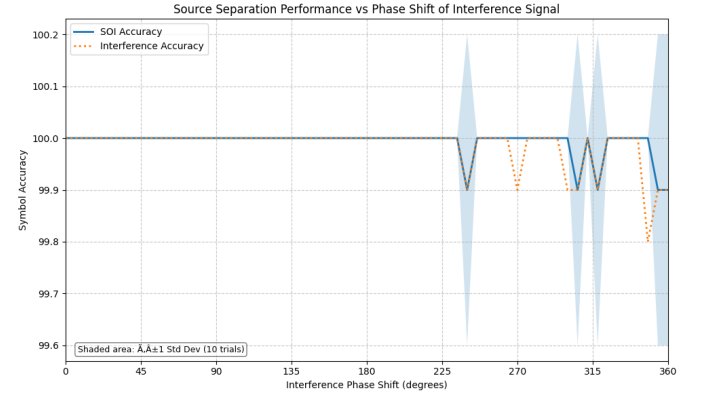


Fig. 15: Dual-channel source separation performance versus phase shift of the interference signal. The solid and dotted lines show mean symbol accuracy for the SOI and interferer, respectively, and the shaded region indicates one standard deviation across repeated trials.

A more pronounced effect appears when the samples-per-symbol parameter is varied at inference time. Table VII and Fig. 16 show that performance is maximized when the inference-time representation matches the training configuration, namely $f_s = 2$. Reducing the temporal resolution to

$f_s = 1$ causes a sharp drop in performance for both sources, while increasing f_s above the training value also degrades performance, though less severely. After the initial drop, the curves largely plateau, indicating that additional temporal resolution does not recover the lost performance.

This result suggests that the dual-channel separator is highly sensitive to representation mismatch. The network was trained on a specific discrete-time structure, and altering that structure at inference time changes the geometry of the input in a way the model does not handle gracefully. As in the single-channel case, higher sampling density is not automatically beneficial. What matters is consistency between the training and inference distributions.

It is also notable that SOI recovery remains consistently stronger than interferer recovery once $f_s \neq 2$. This asymmetry again points to the difficulty of stable source assignment in the dual-channel setting. Under mismatch, the model preserves the more salient source more reliably than the competing source, rather than maintaining equally strong recovery for both.

TABLE VII: Dual-channel source separation performance versus samples-per-symbol f_s . Means are reported as symbol correctness (%), with variance computed across repeated trials.

f_s	SOI Mean	SOI Var	Int Mean	Int Var
1	57.40	11.84	46.40	9.84
2	100.00	0.00	100.00	0.00
10	94.30	5.01	79.90	23.69
20	93.90	3.29	78.60	8.04
30	94.10	2.89	80.60	10.64
40	94.40	5.44	74.40	6.64
50	93.50	6.45	77.60	6.44
60	94.20	2.16	77.30	24.01
70	96.00	2.60	80.40	9.44
80	93.20	9.56	78.70	10.01
90	92.80	12.56	79.50	37.65
100	94.20	2.16	79.50	3.45

Fig. 16 visualizes the same trend. The SOI curve remains above the interferer curve for nearly all $f_s > 2$, confirming that representation mismatch disproportionately harms recovery of the competing source.

Finally, Fig. 17 shows SOI symbol correctness as a function of the timing parameters T_1 and T_2 . The heatmap reveals a clear partition of the parameter plane into a high-performance region and a low-performance region, separated by a sharp boundary that lies close to a diagonal transition. This indicates that the dual-channel model is highly sensitive to the relative timing structure of the SOI and interferer. Unlike the phase-shift experiment, which showed near-invariance, this timing experiment demonstrates that changes in symbol duration substantially alter separability.

This result is consistent with the broader interpretation of the dual-channel problem developed above. The model is not failing because every nuisance transformation is difficult. Rather, it is especially vulnerable to transformations that change the temporal competition between the two sources in a way that is not well represented in training. When timing relationships

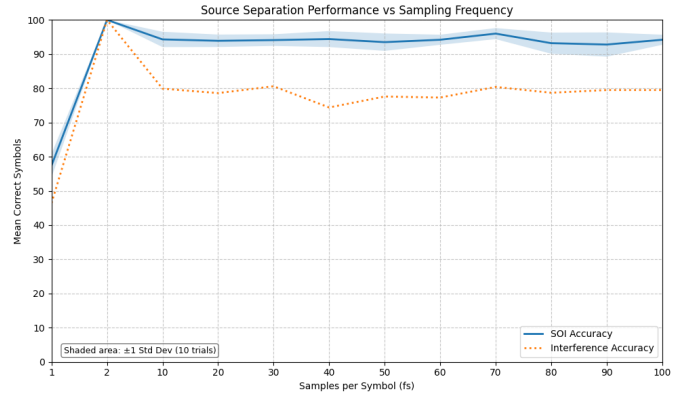


Fig. 16: Dual-channel source separation performance versus samples-per-symbol f_s . The lines show mean symbol accuracy for the SOI and interferer, and the shaded region indicates one standard deviation across repeated trials.

move outside the learned setting, performance drops abruptly rather than gradually.

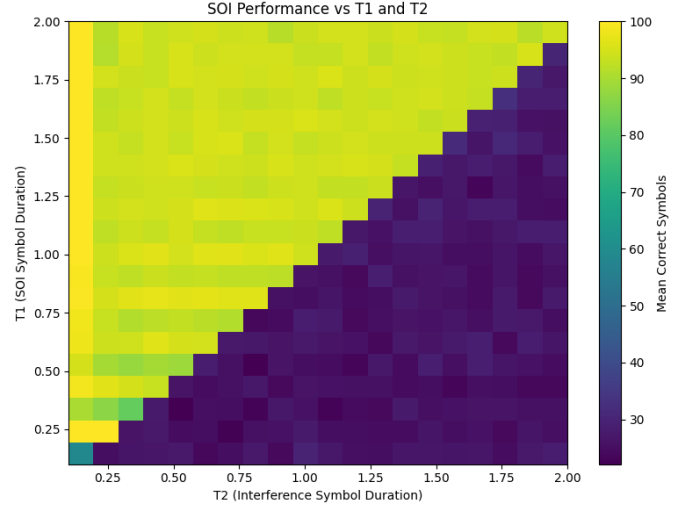


Fig. 17: Dual-channel SOI performance as a function of the timing parameters T_1 and T_2 , reported as mean correct symbols.

Taken together, these additional dual-channel experiments sharpen the interpretation of the main results. The unconstrained dual-channel setting is not merely noisy; it is structurally fragile. The model is comparatively robust to interference phase, but it is much more sensitive to changes in timing structure, samples-per-symbol representation, and joint variations in α and σ^2 . This supports the conclusion that the principal limitation of the unconstrained dual-channel circumstance is not lack of model capacity, but lack of a sufficiently stable and identifiable mixing structure.

C. Quad-Channel Separation Problem

Building upon the insights gained from the symmetry-induced failure modes and the subsequent simplification of the

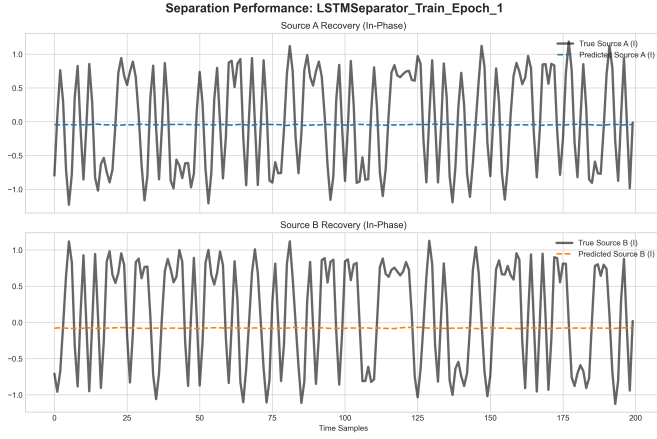


Fig. 18: 1 Epoch

mixing model, the quad-channel experiments ($n_{rx} = 4$) evaluated the architectures using the constrained phase-only mixing matrix. By strictly removing random amplitude scaling while preserving deterministic phase offsets across the four receive antennas, we maintain spatial diversity without introducing pathological ill-conditioning.

The empirical results, summarized in Table VIII, demonstrate a dramatic recovery in model performance. Unlike the dual-channel case, the generalization gap effectively collapsed. The structured mixing environment allowed the learned architectures to fully leverage the spatial information across the four channels.

TABLE VIII: Model comparison including SDR and symbol accuracy.

Model	MSE	SDR (dB)	Train Acc	Val Acc
FastICA	1.20911	-3.71	0.3701	0.3679
Hybrid	0.00004	40.59	1.0000	1.0000
LSTM	0.00001	48.85	1.0000	1.0000
IQ_CNN	0.00016	35.01	1.0000	1.0000
HTDemucs	0.39917	0.98	0.4548	0.4535

The FastICA baseline remained near random guessing (roughly 37% symbol accuracy), underscoring that classical statistical independence alone is insufficient for this heavily structured QPSK mixture. Conversely, the LSTM and Hybrid models achieved perfect symbol recovery (1.000) and exceptional waveform reconstruction.

The stabilization of the learning process is visually evident in the temporal separation progression. Fig. 18 contrasts the LSTM's waveform reconstruction at Epoch 1 versus Epoch 20. Initially, the network struggles to untangle the in-phase and quadrature trajectories, but by Epoch 20, the separated estimations tightly track the ground truth (see Fig. 19).

Progression of LSTM waveform reconstruction in the quad-channel setting, demonstrating rapid convergence onto the true signal trajectories.

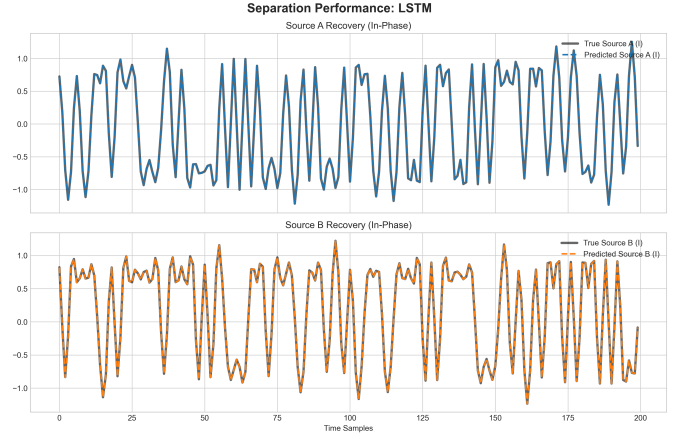


Fig. 19: Final converged LSTM separation for the quad-channel problem, showcasing near-perfect alignment with the ground-truth QPSK waveform.

The HTDemucs architecture also benefited from the phase-only constraint, though its convergence remained distinctly slower than the simpler models. Figures 20 and 21, respectively, illustrate the trajectory from Epoch 1 to Epoch 20. While significant noise suppression is achieved by the final epoch, HTDemucs still exhibits residual variance not seen in the LSTM outputs.

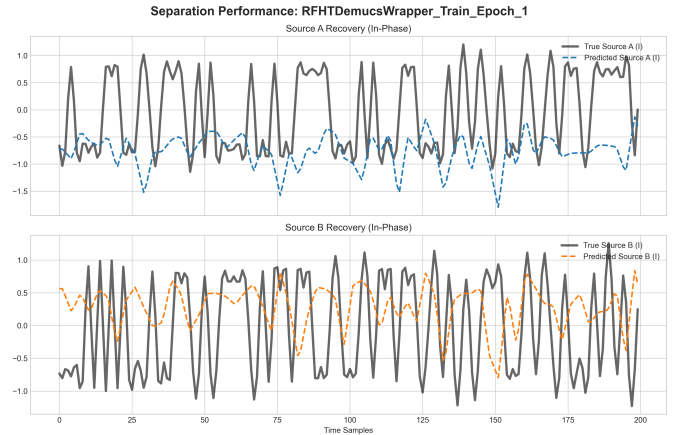


Fig. 20: Epoch 1

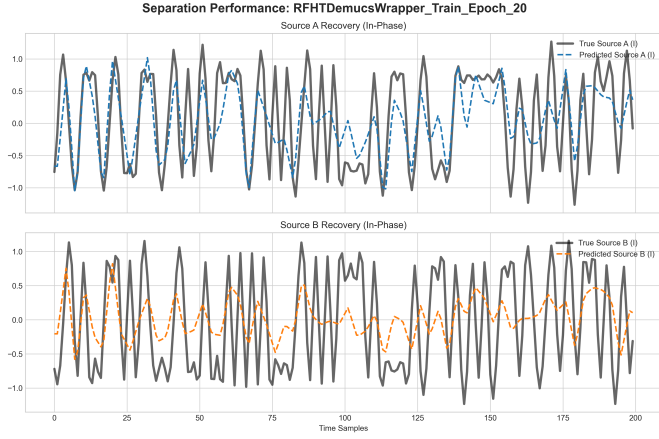


Fig. 21: Epoch 20

HTDemucs progression in the quad-channel setting. Despite improvements, it continues to show higher residual distortion compared to the target-specific RF models. Finally, the global training dynamics for the quad-channel scenario (plotted in Appendix B, Fig. 32 and Fig. 33) completely lack the overfitting divergence observed in the unconstrained dual-channel experiments. This confirms that the phase-only constraint successfully restores waveform identifiability, allowing standard MSE objectives to discover robust separation manifolds.

To further characterize the quad-channel scenario, we performed additional experiments over interference strength α , noise variance σ^2 , interference phase shift, sampling frequency, and relative source timing. These experiments serve two purposes. First, they test whether the strong performance of the constrained four-channel setting is stable under broader perturbations. Second, they clarify which aspects of the recovered performance are due to genuine source separability and which are due to alignment between the training and inference distributions.

Fig. 22 shows the joint sweep over α and σ^2 , reported as mean incorrect symbols for the SOI and interferer. In contrast to the dual-channel case, the quad-channel setting exhibits a much more stable error surface. At $\sigma^2 = 0$, performance is nearly perfect over a wide range of α , and even as noise increases, the degradation is gradual rather than catastrophic. The dominant trend is monotonic in σ^2 : higher noise leads to more symbol errors, but the model continues to recover substantially more structure than would be expected from random guessing. This provides additional evidence that the phase-only constrained mixing matrix restores a robust and learnable separation geometry.

There is still some asymmetry across α . Performance is strongest near moderate interference strengths and weakens somewhat at the extremes, particularly when one source dominates the other. However, this effect is much milder than in the dual-channel setting. The main conclusion from Fig. 22 is therefore that the quad-channel model is no longer dominated by instability in source assignment. Once sufficient spatial structure is present, noise becomes the primary source of

degradation.

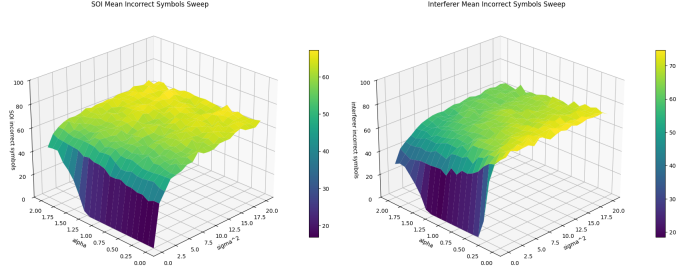


Fig. 22: Mean incorrect symbols in the quad-channel setting as a function of interference strength α and noise variance σ^2 , shown separately for the SOI and the interferer.

The phase-shift experiment reinforces this conclusion. Table IX and Fig. 23 show that symbol recovery remains exactly 100% across the full range of tested interference phase shifts. The mean curves are flat, and the variance is identically zero. This indicates that, in the quad-channel setting, interference phase is not a meaningful failure mode under the constrained mixing model. Unlike the single- and dual-channel cases, the separator has enough spatial information to absorb phase variation without any measurable loss in performance.

TABLE IX: Quad-channel source separation performance versus interference phase shift. Means are reported as symbol correctness (%), with variance computed across repeated trials.

Phase (deg)	SOI Mean	SOI Var	Int Mean	Int Var
0	100.00	0.00	100.00	0.00
30	100.00	0.00	100.00	0.00
60	100.00	0.00	100.00	0.00
90	100.00	0.00	100.00	0.00
120	100.00	0.00	100.00	0.00
150	100.00	0.00	100.00	0.00
180	100.00	0.00	100.00	0.00
210	100.00	0.00	100.00	0.00
240	100.00	0.00	100.00	0.00
270	100.00	0.00	100.00	0.00
300	100.00	0.00	100.00	0.00
330	100.00	0.00	100.00	0.00
360	100.00	0.00	100.00	0.00

Fig. 23 visualizes the same result. Since the curves are constant and the shaded region collapses to zero, the quad-channel separator can be regarded as phase-invariant over this experiment. This is consistent with the broader interpretation of the four-channel system: once the mixture is sufficiently constrained and multiple structured observations are available, phase perturbations alone do not destroy identifiability.

A small but informative degradation appears when the samples-per-symbol parameter is varied at inference time. Table X and Fig. 24 show that the best performance again occurs at the training configuration $f_s = 2$, but unlike the dual-channel case, the loss from distribution mismatch is modest. Increasing f_s above the training value leaves SOI recovery essentially perfect and only slightly reduces interferer recovery, which remains above 98% throughout the sweep. The main

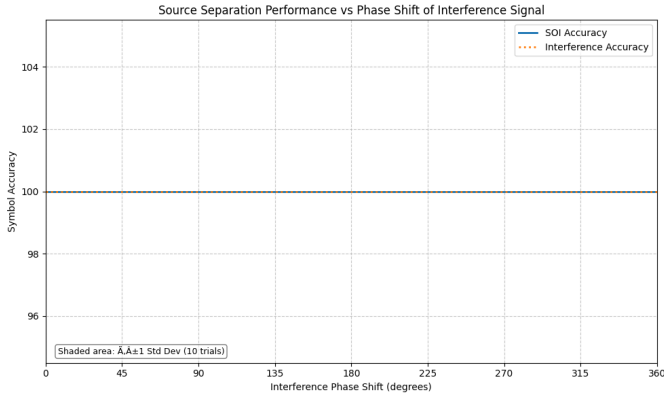


Fig. 23: Quad-channel source separation performance versus phase shift of the interference signal. The solid and dotted lines show mean symbol accuracy for the SOI and interferer, respectively, and the shaded region indicates one standard deviation across repeated trials.

exception is $f_s = 1$, where both sources degrade sharply. This indicates that undersampling removes information that the model genuinely needs, whereas oversampling mostly preserves recoverable structure.

Thus, in the quad-channel setting, the model is still sensitive to representation mismatch, but much less so than in the lower-dimensional cases. The drop from $f_s = 2$ to $f_s = 1$ confirms that temporal resolution still matters. However, once the representation exceeds the training resolution, performance plateaus near the ceiling. This suggests that the quad-channel separator is no longer relying on a fragile coincidence between training and inference grids. Instead, it has learned a more robust mapping that continues to function well under moderate sampling mismatch.

TABLE X: Quad-channel source separation performance versus samples-per-symbol f_s . Means are reported as symbol correctness (%), with variance computed across repeated trials.

f_s	SOI Mean	SOI Var	Int Mean	Int Var
1	68.90	11.49	64.80	22.36
2	100.00	0.00	100.00	0.00
10	99.90	0.09	99.80	0.16
20	100.00	0.00	99.20	0.56
30	100.00	0.00	98.70	1.01
40	100.00	0.00	99.30	0.41
50	100.00	0.00	99.30	1.21
60	100.00	0.00	99.20	1.36
70	100.00	0.00	99.20	0.36
80	100.00	0.00	99.00	0.60
90	99.90	0.09	98.90	0.49
100	100.00	0.00	99.10	0.49

Fig. 24 makes this trend visually clear. The SOI curve is nearly flat at 100% for $f_s \geq 2$, while the interferer curve remains close to perfect with only minor variation. Compared with the dual-channel case, the quad-channel system is therefore substantially more tolerant to changes in temporal resolution.

Finally, Fig. 25 reveals a strong asymmetry when the symbol

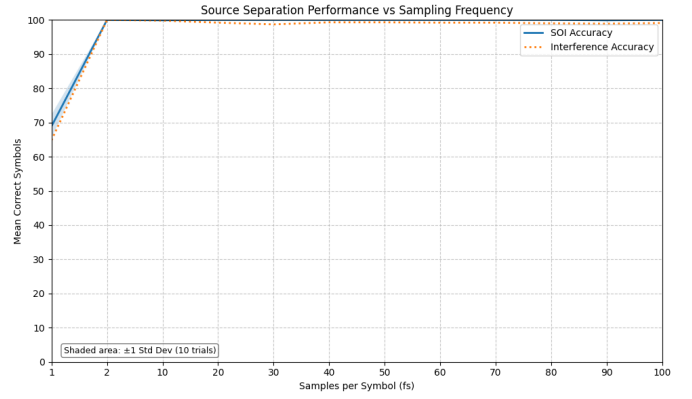


Fig. 24: Quad-channel source separation performance versus samples-per-symbol f_s . The lines show mean symbol accuracy for the SOI and interferer, and the shaded region indicates one standard deviation across repeated trials.

durations T_1 and T_2 are varied. Performance remains very high whenever the SOI has the larger symbol duration, but drops sharply when the interferer has the larger symbol duration. This boundary is visually clean and indicates that the model is exploiting temporal scale differences in a highly systematic way.

The most plausible interpretation is not that higher sampling density is intrinsically helpful, but that the matched-filter and learned temporal features favor the source whose symbol duration is better aligned with the recovery objective. When the SOI has the larger symbol duration, the interferer behaves as a relatively higher-frequency component and is easier to suppress. When the interferer has the larger symbol duration, it behaves as a lower-frequency contaminant that is not removed as effectively by the SOI-oriented processing. In this setting, performance degrades even though the number of receive channels remains unchanged.

This is an important qualification to the otherwise strong quad-channel results. The phase-only constrained mixing matrix restores discernibility and eliminates most of the instability seen in lower-dimensional settings, but the learned separator is still exploiting temporal scale differences in the source structure. Thus, some of its success is due not only to spatial diversity, but also to favorable timing relationships between the sources.

Taken together, these additional experiments sharpen the interpretation of the quad-channel setting. The four-channel phase-only model is substantially more robust than the dual-channel unconstrained model (sophisticated mixing matrix): performance degrades smoothly under noise, remains invariant to interference phase, and is only mildly affected by moderate sampling-frequency mismatch. At the same time, the T_1 - T_2 experiment shows that the model is still leveraging temporal structure in a nontrivial way. Thus, the quad-channel setting should be understood not as a fully general solution to RF source separation, but as a highly successful identifiable scenario in which spatial diversity and structured timing together enable near-perfect recovery.

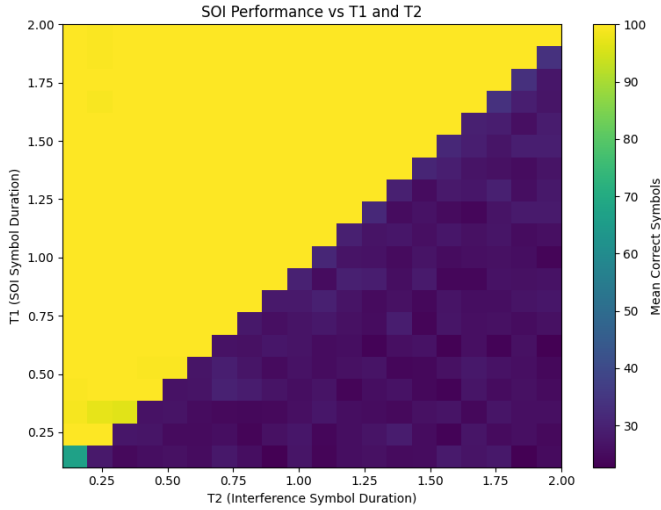


Fig. 25: Quad-channel SOI performance as a function of the timing parameters T_1 and T_2 , reported as mean correct symbols.

VI. CONCLUSION

This work investigated supervised RF source separation under a synthetic QPSK mixture model and showed that separator performance depends fundamentally on the interaction between source identifiability, channel structure, and evaluation metric. In the single-channel setting, the learned models established a strong baseline under moderate interference and low noise, with Hybrid, LSTM, and IQ_CNN all achieving near-perfect symbol recovery and very low waveform error, while HTDemucs converged much more slowly and remained substantially less competitive under the same training budget. The single-channel experiments also revealed two distinct failure mechanisms. The first is symmetry-induced ill-posedness near $\alpha = 1$, where the SOI and interferer have equal strength and are drawn from the same distribution. In that scenario, the mixture becomes permutation-invariant, the inverse problem ceases to be identifiable, and MSE training favors a symmetric averaging solution rather than true source assignment. Empirically, this produced the sharp collapse in both waveform reconstruction and symbol accuracy observed in the α -sweep, together with the characteristic MSE plateau near 0.25. The second failure mechanism is stochastic degradation under increasing noise power. Unlike the abrupt singularity at $\alpha = 1$, the noise sweeps produced a gradual and monotone decline in symbol recovery, indicating that additive noise perturbs an otherwise learnable mapping rather than destroying identifiability itself. Additional single-channel experiments further clarified that the learned separator is comparatively robust to interference phase shifts, but much more sensitive to inference-time mismatch in temporal sampling and timing structure. In particular, performance was strongest when the samples-per-symbol configuration matched the training distribution, and degraded noticeably when the input representation was undersampled or oversampled relative to training. Taken together, the single-channel results show that learned RF separation can be highly effective even in an

underdetermined setting, but only when the source geometry remains statistically favorable and the inference distribution remains close to the training circumstance.

The dual-channel results showed that adding receivers alone is not sufficient to guarantee reliable separation. Under the original unconstrained mixing model, where each receive channel included random amplitude and phase variation, the learned models exhibited severe overfitting: training accuracy increased rapidly, but validation symbol accuracy remained poor, and the gap between training and validation performance became one of the dominant empirical features of the two-channel scenario. This demonstrated that the central limitation was not model capacity, since the networks could memorize the training mixtures, but rather the absence of a sufficiently stable and identifiable mixing structure across channel realizations. The additional dual-channel sweeps reinforced this interpretation. The joint (α, σ^2) surfaces showed broad regions of elevated symbol error rather than a narrow isolated failure band, indicating that the model was unstable across a wide range of interference and noise conditions. Phase-shift sweeps remained nearly flat, showing that interference phase by itself was not the principal source of dual-channel failure. In contrast, timing-structure experiments and samples-per-symbol sweeps exposed strong sensitivity to representation mismatch and temporal competition between the sources. Inference performance was best when the representation matched training, and degraded substantially when the temporal structure was perturbed, especially for the interferer. These observations support a consistent conclusion: in the unconstrained dual-channel problem, the model does not learn a robust inverse map from mixtures to sources, but instead relies on fragile regularities tied to the training distribution. As a consequence, waveform MSE and symbol correctness become even more decoupled in this circumstance; low waveform error does not imply correct source assignment, and the model can appear to reconstruct the mixture structure while still failing to recover the transmitted information.

The quad-channel experiments demonstrated that the main obstacle in the lower-dimensional settings was indeed discernibility rather than an inherent limitation of deep learning for RF communications. By replacing the unconstrained multichannel mixing model with a phase-only constrained mixing matrix, the experiments preserved spatial diversity while eliminating the pathological conditioning introduced by random amplitude scaling. Under this structured four-channel geometry, the learned models recovered dramatically. The generalization gap effectively disappeared, validation performance became nearly identical to training performance, and the best task-specific models achieved essentially perfect symbol recovery together with extremely low waveform MSE and high SDR. In this scenario, the learned separator no longer behaved like a memorizer of channel realizations, but rather like a stable estimator operating on a well-posed inverse problem. The additional quad-channel experiments strengthened this conclusion. The joint (α, σ^2) sweeps showed broad regions of near-perfect performance with only gradual

degradation under increasing noise, rather than the unstable error surfaces seen in the dual-channel case. Phase-shift sweeps were completely flat, indicating practical phase invariance under the constrained geometry. Sampling-frequency sweeps showed that the quad-channel separator remained strongest at the training configuration, but was much more tolerant to moderate distribution mismatch than either the single- or dual-channel systems; once the representation exceeded the training resolution, performance remained near the ceiling. At the same time, the T_1 - T_2 timing experiments revealed an important qualification: although the constrained four-channel model restores identifiability, the learned separator still exploits temporal scale differences between the SOI and the interferer. Performance remained strongest when the SOI timing was more favorable to the matched-filter and learned temporal features, and deteriorated when the interferer occupied the more easily preserved temporal scale. Thus, the quad-channel setting should not be interpreted as a completely general solution to RF source separation, but rather as a highly successful identifiable operating setting in which structured spatial diversity and favorable temporal relationships allow deep learning models to recover both waveform structure and symbol information with near-perfect reliability.

Across all three circumstances, several broader conclusions emerge for machine learning in RF communications. First, symbol-level evaluation is indispensable. Throughout this work, waveform MSE, SDR, and symbol recovery were related but not equivalent, and the experiments repeatedly showed that a model can look favorable under waveform reconstruction metrics while still failing at the communications objective of correct symbol decoding. Second, source separation in RF is not determined by network architecture alone; it is governed by the geometry of the mixing process. The comparison among single-, dual-, and quad-channel settings shows that the same family of models can move from near-perfect success, to complete instability, and back to robust generalization depending on whether the underlying inverse problem is identifiable. Third, data-generation assumptions matter enormously. Noise level, interference strength, temporal sampling, timing structure, and the form of the channel matrix all change not only the numerical difficulty of the task but the very nature of the function the model is being asked to learn. For RF machine learning systems intended for deployment, this implies that training data must be designed to reflect the invariances and ambiguities of the operational environment rather than simply to provide large volumes of synthetic examples. Finally, the results suggest that deep learning can indeed surpass classical blind source separation methods such as FastICA in structured RF environments, but only when the problem formulation preserves enough information for source identity to be recoverable in the first place. In that sense, the main lesson of this work is not merely that certain architectures perform better than others, but that successful ML-based RF separation requires a principled alignment among channel model, source geometry, training objective, and communications-level evaluation. When that alignment is absent, even strong models fail for structural

reasons; when it is restored, they can achieve extremely high-fidelity recovery and near-perfect symbol correctness.

REFERENCES

- [1] A. Défossez, N. Usunier, L. Bottou, and F. Bach, “Music source separation in the waveform domain,” in *Proc. Int. Soc. Music Inf. Retrieval Conf. (ISMIR)*, 2019.
- [2] Massachusetts Institute of Technology, “Rf challenge: Ai accelerator,” <https://rfchallenge.mit.edu/>, 2026, accessed: Apr. 2026.
- [3] B. Sklar, *Digital Communications: Fundamentals and Applications*, 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR, 2000.
- [4] Facebook AI Research, “Demucs: Music source separation,” <https://github.com/facebookresearch/demucs>, 2026, accessed: Apr. 2026.
- [5] A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
- [6] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.
- [7] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2016.
- [8] Y. Wu and K. He, “Group normalization,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
- [9] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus),” *arXiv preprint arXiv:1606.08415*, 2016.
- [10] A. Défossez, “demucs,” <https://pypi.org/project/demucs/>, 2026, accessed: Apr. 2026.
- [11] N. Takahashi and Y. Mitsufuji, “D3net: Densely connected multi-dilated densenet for music source separation,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2021, pp. 851–855.
- [12] OpenCodePapers, “Music source separation on musdb18 benchmark,” <https://opencodepapers-b7572d.gitlab.io/benchmarks/music-source-separation-on-musdb18.html>, 2023, accessed: Apr. 2026.

VII. APPENDIX

This appendix provides additional figures supporting the main experimental results. While the main text focuses on high-level trends, these figures illustrate detailed behaviors in training dynamics, α -dependent transitions, and noise sensitivity.

APPENDIX A SYSTEM ARCHITECTURES

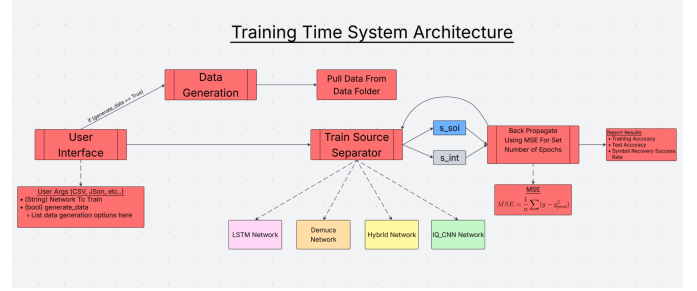


Fig. 26: System architecture for the training phase.

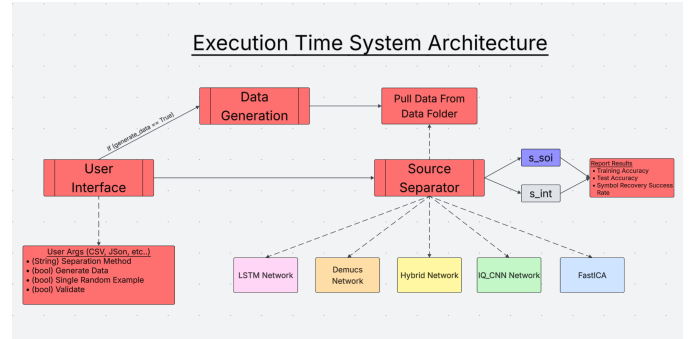


Fig. 27: System architecture for the evaluation and execution phase.

APPENDIX B TRAINING AND VALIDATION CURVES FOR THE NETWORKS IN SINGLE, DUAL, AND QUAD CHANNEL SETTINGS

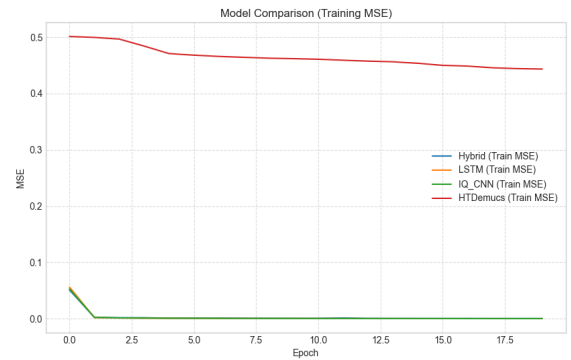


Fig. 28: Single Channel Training Plot

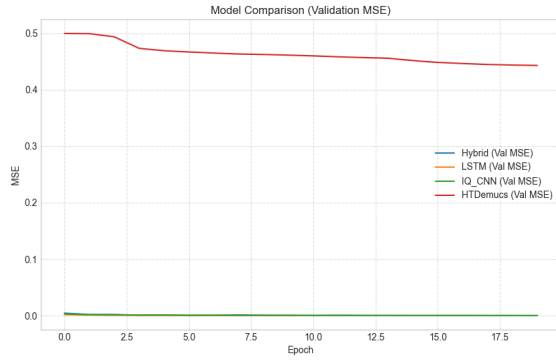


Fig. 29: Single Channel Validation Plot

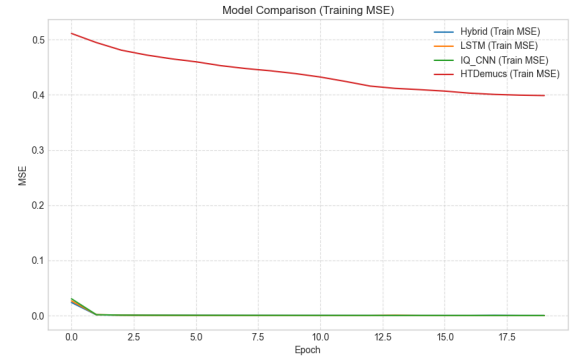


Fig. 32: Quad Channel Training Plot

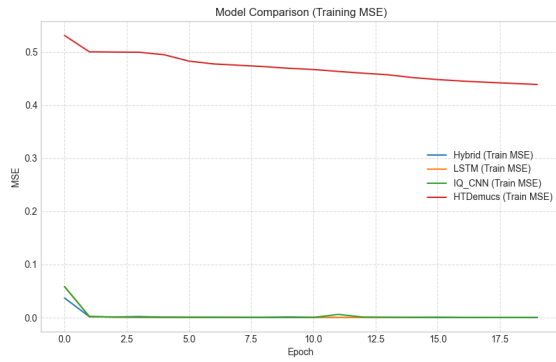


Fig. 30: Dual Channel Training Plot

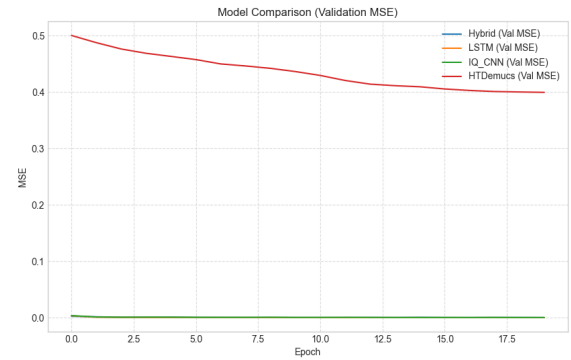


Fig. 33: Quad Channel Validation Plot

APPENDIX C MUSIC SOURCE SEPARATION

Early in the project, the focus was on music source separation (MSS), a task in which a mixed audio signal is decomposed into constituent sources such as vocals, drums, and bass. Modern MSS approaches rely heavily on deep learning, improving significantly over earlier methods based on non-negative matrix factorization and handcrafted features.

Experiments were conducted using the MUSDB18 dataset, which contains 150 fully labeled music tracks with isolated source components. Initial efforts followed the sponsor's recommendation to implement D3Net [11], a convolutional encoder-decoder architecture that combines dilated convolutions, dense skip connections, and multi-scale feature extraction. While D3Net provided a useful baseline, a subsequent literature review indicated that it no longer represents the state of the art.

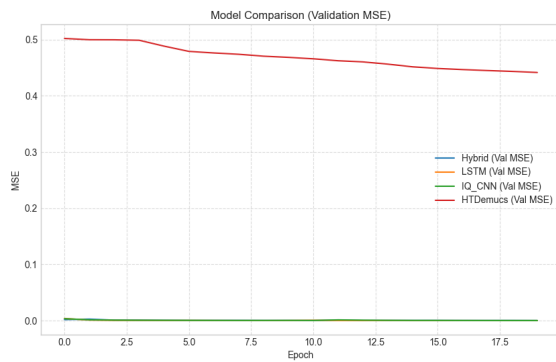


Fig. 31: Dual Channel Validation Plot

Leaderboard								
Click a sortable column header to sort.			Show papers without code					
Paper	Code	SDR (avg)	SDR (vocals)	SDR (drums)	SDR (bass)	SDR (other)	ModelName	ReleaseDate
Hybrid Transformers for Music Source Separation	✓ Link	9.20	9.37	10.83	10.47	6.41	Sparse HT Demucs (fine tuned)	2022-11-15
Hybrid Transformers for Music Source Separation	✓ Link	9.00	9.20	10.08	9.78	6.42	Hybrid Transformer Demucs (f.t.)	2022-11-15
Music Source Separation with Band-split RNN	✓ Link	8.97	10.47	10.15	8.16	7.08	Band-Split RNN (semi-sup.)	2022-09-30
Sound Demixing Challenge 2023 Music Demixing Track Technical Report: TFC-TDF-UNet v3	✓ Link	8.34	9.59	8.44	8.45	6.86	TFC-TDF-UNet (v3)	2023-06-15
Music Source Separation with Band-split RNN	✓ Link	8.23	10.21	8.58	7.51	6.62	Band-Split RNN	2022-09-30
Hybrid Spectrogram and Waveform Source Separation	✓ Link	7.72	8.04	8.58	8.67	5.59	Hybrid Demucs	2021-11-05
KUIELab-MDX-Net: A Two-Stream Neural Network for Music Demixing	✓ Link	7.54	9.00	7.33	7.86	5.95	KUIELab-MDX-Net	2021-11-24
An Efficient Short-Time Discrete Cosine Transform and Attentive MultiResUNet Framework for Music Source Separation	✓ Link	6.81	8.57	7.63	5.88	5.14	Attentive-MultiResUNet	2022-11-14
Music Source Separation in the Waveform Domain	✓ Link	6.79	7.29	7.58	7.60	4.69	DEMUCS (extra)	2019-11-27
CWS-PRResUNet: Music Source Separation with Channel-wise Subband Phase-aware ResUNet	✓ Link	6.77	8.92	6.38	5.93	5.84	CWS-PRResUNet	2021-12-09
D3Net: Densely connected multilayered DenseNet for music source separation	✓ Link	6.68	7.80	7.36	6.20	5.37	D3Net	2020-10-05

Fig. 34: Comparison of MSS model performance on MUSDB18 (adapted from [12]).

This motivated a transition to the Demucs architecture, shown in Fig. 35, which operates directly in the time domain and has demonstrated strong performance on MSS benchmarks. The model was successfully implemented and used to achieve high-quality source separation.

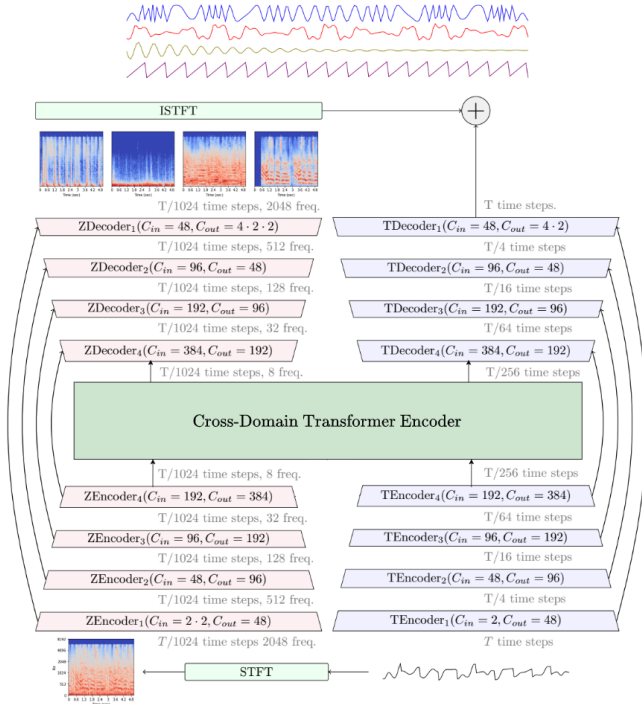


Fig. 35: Hybrid Transformer Demucs architecture (Fig. from [4]).

The success of Demucs in time-domain audio separation

motivated its application to RF source separation. Although MSS and RF differ in signal structure and evaluation criteria, both involve recovering constituent signals from mixtures, making this transfer a natural starting point for subsequent experiments.

APPENDIX D MIT INTERFERENCE CHALLENGE

Following the MSS phase, the project shifted to RF source separation. As an initial step, the MIT Interference Challenge dataset [2] was explored. This dataset, however, is not structured for supervised learning, as ground-truth separated signals are not directly available.

Attempts to apply supervised learning methods were therefore unsuccessful, motivating a transition to unsupervised (blind) separation techniques. One approach explored was per-frame DeepICA, which approximates FastICA using a neural network trained with decorrelation and non-Gaussianity objectives. While this method produced modest improvements in bit error rate, it was not sufficient to recover complete symbol sequences.

To better capture time-dependent mixing effects, convolutional layers were incorporated into the model. This modification yielded incremental improvements but still did not achieve reliable full-frame decoding.

Although these approaches did not produce competitive final results, they clarified key challenges in RF source separation, including the role of temporal structure and the difficulty of learning without explicit supervision. These insights informed the design of the synthetic data pipeline and the learned models used in the main body of this work.

APPENDIX E ADDITIONAL RF SWEEP FIGURES

This appendix collects the most informative extra figures from the single-channel experiments. The main text emphasizes the highest-level findings, while the appendix provides more detailed evidence for the training behavior, the α -sweep transition, the noise-sweep transition, and the qualitative differences between near-perfect and degraded waveform recovery.

A. HTDemucs Training and Qualitative Results

HTDemucs learned much more slowly than Hybrid, LSTM, and IQ-CNN in the single-channel setting. Even after its longer run, its waveform reconstruction remained noticeably worse than the smaller learned models. This slower convergence, together with weaker final waveform quality, is why HTDemucs was omitted from the α and noise sweeps.

B. Additional Alpha-Sweep Figures

The full α sweep showed that the most difficult point is the equal-strength mixture near $\alpha = 1.0$. This is visible not only in waveform MSE and SDR, but also in symbol accuracy and in the validation-curve dynamics near the transition region.

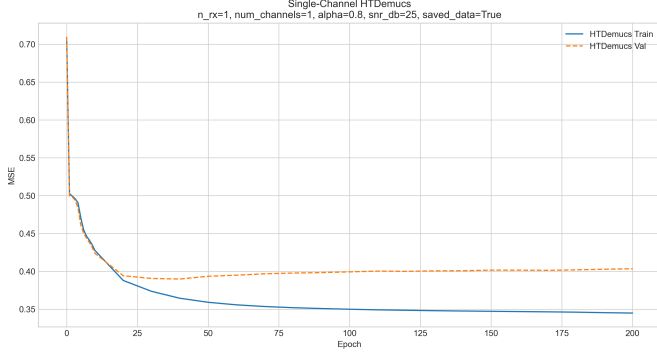


Fig. 36: HTDemucs single-channel training and validation curves. Compared with the other learned models, HTDemucs improved much more slowly and did not approach the same low-error setting within the short comparison budget.

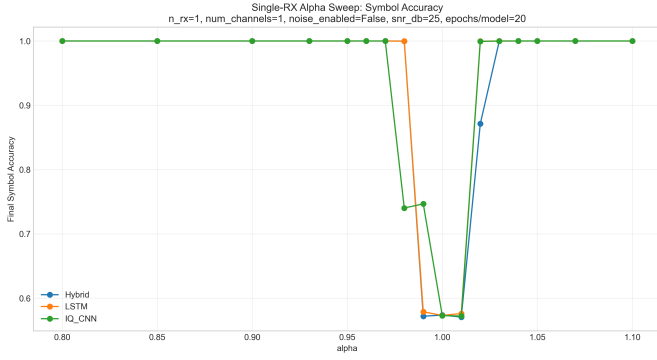


Fig. 37: Full single-channel α sweep measured by symbol accuracy. The narrow failure region is concentrated near the equal-strength mixture around $\alpha = 1.0$.

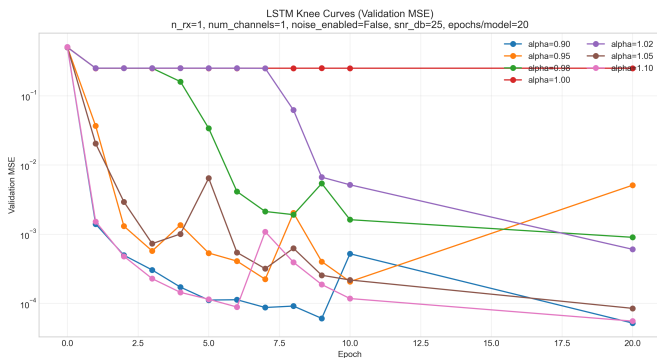


Fig. 38: LSTM validation curves for selected α values near the failure region. The abrupt degradation near $\alpha = 1.0$ is consistent with the symmetry-based interpretation discussed in the main text.

C. Additional Noise-Sweep Figures

Unlike the α sweep, the noise sweep did not produce a narrow catastrophic region. Instead, waveform and symbol performance degraded gradually as the noise variance increased.

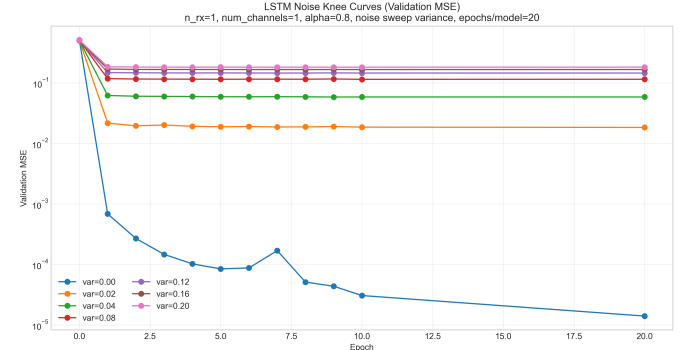
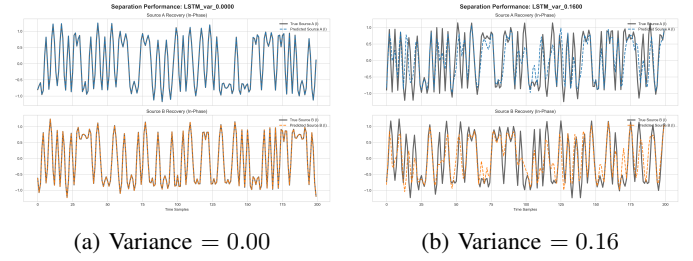


Fig. 39: LSTM validation curves for selected noise variances. The degradation is smooth as variance increases, in contrast to the sharper α -driven failure near the equal-strength case.



(a) Variance = 0.00

(b) Variance = 0.16

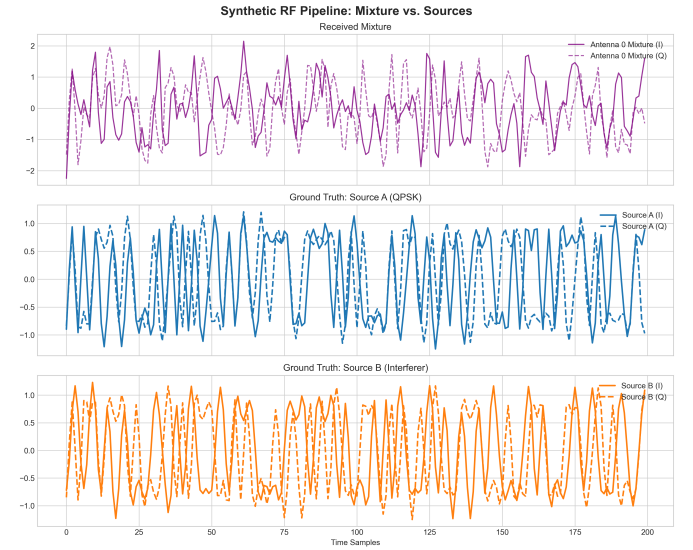


Fig. 41: Example data-pipeline view for a high-noise case. This Figure illustrates the visual difference between the clean source structure and the observed mixture at a difficult noise level.

TABLE XI: Full single-channel α sweep results.

α	Model	Best Epoch	Best Val MSE	Final Val MSE	Final SDR (dB)	Final Symbol Acc.	Stop Epoch
0.80	Hybrid	17	0.000165	0.000165	34.81	1.0000	20
0.80	LSTM	17	0.000019	0.000019	44.22	1.0000	20
0.80	IQ_CNN	19	0.000040	0.000040	41.06	1.0000	20
0.85	Hybrid	18	0.000087	0.000087	37.62	1.0000	20
0.85	LSTM	15	0.000017	0.000017	44.62	1.0000	20
0.85	IQ_CNN	16	0.000053	0.000053	39.79	1.0000	20
0.90	Hybrid	20	0.000109	0.000109	36.61	1.0000	20
0.90	LSTM	20	0.000052	0.000052	39.83	1.0000	20
0.90	IQ_CNN	10	0.000135	0.000135	35.73	1.0000	20
0.93	Hybrid	20	0.000106	0.000106	36.73	1.0000	20
0.93	LSTM	20	0.000063	0.000063	38.99	1.0000	20
0.93	IQ_CNN	20	0.000112	0.000112	36.59	1.0000	20
0.95	Hybrid	17	0.000115	0.000115	36.40	1.0000	20
0.95	LSTM	16	0.000089	0.000089	37.64	1.0000	20
0.95	IQ_CNN	19	0.000129	0.000129	35.94	1.0000	20
0.96	Hybrid	19	0.000096	0.000096	37.16	1.0000	20
0.96	LSTM	20	0.000070	0.000070	38.71	1.0000	20
0.96	IQ_CNN	20	0.000376	0.000376	31.31	1.0000	20
0.97	Hybrid	19	0.000183	0.000183	34.39	1.0000	20
0.97	LSTM	19	0.000207	0.000207	34.17	1.0000	20
0.97	IQ_CNN	20	0.000617	0.000617	29.21	1.0000	20
0.98	Hybrid	20	0.000797	0.000797	28.17	0.9998	20
0.98	LSTM	16	0.000696	0.000696	28.88	0.9996	20
0.98	IQ_CNN	20	0.125624	0.125624	6.00	0.7403	20
0.99	Hybrid	9	0.249625	0.249625	3.02	0.5721	20
0.99	LSTM	15	0.249553	0.249553	3.02	0.5789	20
0.99	IQ_CNN	20	0.128958	0.128958	5.89	0.7467	20
1.00	Hybrid	19	0.249567	0.249567	3.02	0.5738	20
1.00	LSTM	6	0.249504	0.249504	3.02	0.5733	20
1.00	IQ_CNN	13	0.249526	0.249526	3.02	0.5730	20
1.01	Hybrid	16	0.249981	0.249981	3.01	0.5707	20
1.01	LSTM	16	0.249877	0.249877	3.02	0.5762	20
1.01	IQ_CNN	15	0.249922	0.249922	3.01	0.5723	20
1.02	Hybrid	20	0.089617	0.089617	7.48	0.8713	20
1.02	LSTM	20	0.000605	0.000605	29.62	0.9996	20
1.02	IQ_CNN	20	0.022320	0.022320	13.52	0.9993	20
1.03	Hybrid	19	0.000113	0.000113	36.47	1.0000	20
1.03	LSTM	19	0.000134	0.000134	35.91	1.0000	20
1.03	IQ_CNN	20	0.000611	0.000611	29.36	0.9998	20
1.04	Hybrid	19	0.000151	0.000151	35.21	1.0000	20
1.04	LSTM	20	0.000051	0.000051	39.98	1.0000	20
1.04	IQ_CNN	12	0.000487	0.000487	30.21	1.0000	20
1.05	Hybrid	15	0.000143	0.000143	35.45	1.0000	20
1.05	LSTM	20	0.000085	0.000085	37.74	1.0000	20
1.05	IQ_CNN	18	0.000154	0.000154	35.21	1.0000	20
1.07	Hybrid	20	0.000111	0.000111	36.55	1.0000	20
1.07	LSTM	20	0.000044	0.000044	40.61	1.0000	20
1.07	IQ_CNN	20	0.000140	0.000140	35.59	1.0000	20
1.10	Hybrid	20	0.000129	0.000129	35.88	1.0000	20
1.10	LSTM	19	0.000041	0.000041	40.84	1.0000	20
1.10	IQ_CNN	15	0.000085	0.000085	37.77	1.0000	20