uc3m | Universidad Carlos III de Madrid

Master's Degree in Machine Learning for Health
2024-2025

Master Thesis

"On the Generalization of Deep Neural Network-Based Interference Rejection to Unseen Signal Constellations"

Mario Golbano Corzo

Alejandro Lancho Serrano

Leganés, September 2025

On the Generalization of Deep Neural Network-Based Interference Rejection to Unseen Signal Constellations

Mario Golbano Corzo
Author

Universidad Carlos III de Madrid
100504162@alumnos.uc3m.es

Alejandro Lancho Serrano
Supervisor

Signal Theory and Communications Departament
Universidad Carlos III de Madrid
alancho@ing.uc3m.es

***Abstract*—The use of machine learning for interference rejection in wireless communications has gained increasing attention in recent years. While classical methods such as matched filtering and linear MMSE estimation remain strong baselines, there is still no universally reliable data-driven approach for separating superimposed RF signals. In this work, we investigate the application of deep neural networks to the problem of single-channel RF source separation. Building on architectures originally developed for audio processing, we adapt the HT-Demucs model to the wireless domain and evaluate its effectiveness in recovering a signal of interest from controlled mixtures, assessing performance in terms of both signal quality and bit error rate (BER). A key enabler of our study is a purpose-built dataset and generation framework tailored to this application. It modulates video-derived bitstreams into OFDM waveforms with diverse constellation orders and composes mixtures with technology-specific interferers under controlled SINR schedules and other variability parameters. This design provides both diversity for training and principled, reproducible evaluation while preserving a semantic layer for downstream analysis. Finally, we examine the ability of learned models to handle unseen modulation constellations, providing insights into their generalization capabilities. The results highlight the potential of data-driven approaches as a complement to classical signal processing, while underscoring the challenges that remain in achieving robust, modulation-independent interference rejection.**

***Index Terms*—Interference mitigation, source separation, data-driven methods, wireless communications, generalization**

I. Introduction

WIRELESS communications are central to modern society, yet increasingly exposed to interference as spectrum demand grows and multiple signals overlap in time and frequency. Recovering the intended message under such conditions remains a key challenge for receiver design.

Classical approaches such as matched filtering and linear MMSE estimation [1] provide solid baselines, but rely on accurate channel knowledge and statistical assumptions that may not hold in dynamic wireless environments. This has motivated the rise of data-driven methods, where neural networks learn separation strategies directly from examples, bypassing the need for explicit parametric models.

Crucially, the goal of communication systems is not always the exact recovery of the transmitted waveform, but rather the preservation of the conveyed information. This perspective, formalized in the paradigm of semantic communications, shifts the emphasis from signal fidelity to message intelligibility. In this work, video transmissions are employed as a practical testbed: even when the recovered video is imperfect, its content may remain understandable. This enables both quantitative evaluation, through metrics such as BER after demodulation, and qualitative assessment of how well machine learning–based interference rejection preserves meaning [2].

A. Previous Work

Traditional interference rejection techniques in wireless communications have relied on strong theoretical foundations and well-known signal-processing tools. Methods like matched filtering and linear MMSE estimation [1] leverage prior information, such as channel models or statistical properties of interference, to remove unwanted signals. While effective under their assumptions, these techniques can struggle in highly dynamic or heterogeneous environments where such assumptions break down.

More advanced model-based approaches, such as maximum likelihood sequence estimation or particle filtering, can in principle improve performance but rely heavily on detailed signal models. When these are inaccurate, performance degrades [3] [4].

The paradigm of data-driven interference rejection emerged as an alternative, motivated by the limitations of classical methods. To address these limitations, recent work has explored deep neural networks (DNNs) that learn separation directly from datasets rather than relying on parametric assumptions. The RF Challenge organized by MIT at ICASSP 2024 established a benchmark framework for testing single-channel RF interference rejection algorithms [5]. Building on this benchmark, in [6] it was introduced some baseline architectures such as UNet and WaveNet, and demonstrated that these learning-based methods can achieve substantially

lower BER than classical techniques like matched filtering or linear MMSE estimation.

The UNet model, originally proposed for biomedical image segmentation [7], follows an encoder–decoder structure where the encoder reduces temporal resolution while extracting features, and the decoder reconstructs the signal by upsampling. With skip connections preserving fine-grained details, UNet has proven effective in source separation tasks, including RF interference rejection in the RF Challenge, where it achieved lower BER than classical baselines. It has also been used to test data-driven frameworks for wireless communications, serving as a baseline to validate interference rejection and laying the groundwork for more advanced architectures [6].

RF signals differ substantially from natural signals like images or audio, as they are digitally generated, symbol-based, and often overlap in time and frequency with dependencies across multiple scales. Many machine learning approaches so far have focused on analyzing them in the time domain, directly from I/Q samples. However, in communication systems the actual information is encoded and recovered in the frequency domain, where key structures of the signal become more apparent. This suggests that restricting analysis to one domain may overlook useful information. For this reason, we propose an approach that processes RF signals jointly in time and frequency, aiming to exploit the complementary features of both representations to achieve more effective interference rejection, and we ground the evaluation on the tailored dataset described above.

Beyond algorithms, *dataset design* has become central to progress. Public benchmarks have catalyzed research, but they often fix specific conditions and evaluation protocols. In this work we complement existing resources with a *purpose-built dataset and generation framework* [8] tailored to interference mitigation: video-derived bitstreams are modulated into OFDM waveforms with diverse constellation orders, mixtures are composed with technology-specific interferers (e.g., Bluetooth/Wi-Fi), and variability is controlled through SINR schedules and related parameters. This provides training diversity and, crucially, enables *semantic* evaluation via demodulation (BER) and recovered-video quality under reproducible settings.

B. Contributions

The contributions of this work can be summarized as follows:

- *Joint time–frequency modeling for RF separation* We adapt the HT-Demucs architecture, originally developed for music source separation, to the wireless domain. By combining time-domain I/Q processing with frequency-domain spectrogram representations, the model captures complementary structures of RF signals that would be overlooked if analyzed in a single domain.
- *Generalization across modulation schemes* Our experiments explicitly evaluate how learned models behave when exposed to signal constellations not seen during training. This provides insights into the robustness and limitations of data-driven interference rejection, an important step toward modulation-agnostic receivers.
- *Applicability to interference mitigation and source separation* Beyond academic benchmarks, the proposed approach addresses a practical challenge in wireless communications: separating a signal of interest from overlapping transmissions. This makes the method directly relevant for improving receiver performance in interference-limited environments.
- *Dataset contribution and flexibility* A central contribution of this work is the use of a controlled dataset that extends well beyond the OFDM case studied here. The dataset generation framework employed is designed to flexibly support diverse wireless technologies and modulation formats under realistic interference conditions, enabling the project to be extrapolated to a broad range of experiments on interference mitigation in wireless communications.
- *Semantic perspective on wireless communications* By incorporating video-based transmissions, our evaluation goes beyond waveform-level accuracy. We analyze bit error rate (BER) after demodulation as a proxy for how much of the transmitted information can still be understood, aligning our study with the emerging paradigm of semantic communications.

II. Problem Statement

We consider the task of recovering a *signal of interest* (SoI) from a received mixture that also contains an interfering signal and additive noise. Both signals overlap in time and frequency. In this setting, the challenge is not only to isolate the SoI waveform but also to ensure that it can be successfully demodulated to recover the underlying information. For this purpose, we model the communication as the point-to-point, single-channel baseband transmission of an encoded and modulated bitstream, where the interference and some noise is added (shown in shown in Fig 1).

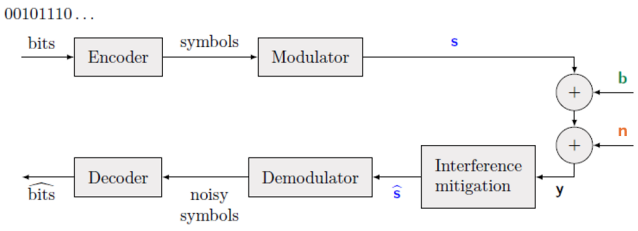


Fig. 1. Communication scheme with a dedicated building block for interference mitigation [6].

This can also be represented by

$$y = s + b + n \in \mathbb{C}^{Nx1} \tag{1}$$

This simplified signal model allows us to focus on the problem of interference mitigation, which is the main objective of the project, not considering other concerns for the signal.

The goal is to design a method that, given y, produces an estimate $\hat{s}$ of the SoI that allows accurate demodulation. This

formulation connects the task to both *source separation* and *interference rejection*, while also highlighting the importance of performance metrics at the symbol or bit level rather than waveform fidelity alone.

A. Signal Datasets

In order to train and evaluate data-driven interference mitigation techniques, it is necessary to rely on synthetic signal datasets that can reproduce realistic wireless communication scenarios while remaining fully controllable. In our case, both the signal of interest (SoI) and the interfering signals are generated using a dedicated framework [8], which allows us to have the transmitted bitstreams derived from encoded video sequences. This adds a semantic layer to the dataset: the machine learning models are not only trained on symbol-level structures or on the spectogram, but can also be evaluated on their ability to preserve the intelligibility of the underlying video content after demodulation. All this provides complete knowledge of the generation process, allowing us to focus on the recovery of the SoI. Since we restrict ourselves to two-signal mixtures, the source separation task can be formulated as learning to predict the SoI; in principle, subtracting this estimate from the received mixture would yield the interferer. Nonetheless, in this work we focus specifically on interference mitigation, i.e., recovering the SoI with sufficient quality for demodulation and decoding.

We concentrate on multi-carrier transmissions, given their prevalence in modern wireless systems. In particular, we adopt orthogonal frequency-division multiplexing (OFDM), a widely used technology that underpins standards such as WiFi, LTE, and 5G. An OFDM signal is formed by modulating information symbols onto multiple orthogonal subcarriers, each carrying a symbol from a chosen constellation. In our dataset, OFDM signals are generated using different modulation orders, namely QPSK, 64-QAM, and 1024-QAM, covering a wide range of spectral efficiencies and robustness levels. This variety is key to testing the ability of data-driven models to generalize across distinct constellations [9].

To increase realism and challenge the interference rejection task, we consider interference sources that are representative of practical environments. Most notably, Bluetooth transmissions are included, as they operate in the same unlicensed 2.4 GHz band as WiFi and therefore constitute a common source of interference in practice [10]. In addition, we incorporate other WiFi transmissions as interfering signals for the testset. This interfering WiFi transmissions, while being different signals from the used SoIs, are still based on this OFDM modulation technology [11], introducing further complexity into the interference scenario. This mixture of controlled yet realistic conditions makes the dataset a suitable testbed for evaluating data-driven interference mitigation approaches.

III. Data-driven Methods

This section introduces the neural architectures employed in our study for data-driven interference mitigation. Our choice of models is informed both by prior work in RF signal separation and by insights gained from the audio source separation domain. While classical receivers are typically designed around matched filtering and MMSE estimation [1], these rely on accurate models of the channel and interference, which are rarely available in practice. Data-driven methods, by contrast, aim to learn directly from examples of signal mixtures and their corresponding clean targets, thereby capturing statistical structures that are otherwise analytically intractable.

Following the approach established in related works, we cast the interference rejection problem as a multivariate regression task. The input to the neural network is the complex-valued baseband mixture y, represented as two channels corresponding to its in-phase (I) and quadrature (Q) components. The output is an estimate $\hat{s}$ of the signal of interest (SoI), with the same dimensionality as the input. The model is thus trained end-to-end to minimize the discrepancy between $\hat{s}$ and the ground truth s, measured using loss functions aligned with communication objectives [6].

A key design challenge lies in tailoring deep learning architectures to the RF domain. Unlike images or audio, RF signals are generated by digital modulation schemes and exhibit both short- and long-range temporal dependencies. Moreover, their spectral structure is tightly coupled with the underlying modulation and multiplexing techniques, such as OFDM. These characteristics motivate the exploration of architectures capable of extracting features across multiple temporal scales while preserving fine-grained details of the waveform. In this work, we evaluate a modified time–frequency architecture based on HT-Demucs [12] [13].

A. HT-Demucs: Hybrid Time-Frequency Separation with Cross-Domain Attention

HT-Demucs (*Hybrid Transformer Demucs*) is a hybrid source separation architecture that runs two UNets in parallel; one in the time domain on raw baseband I/Q and one in the frequency domain on complex STFTs; and couples them through a cross-domain Transformer at the bottleneck (Fig 2). The motivation is that relevant RF structure is complementary across domains: local waveform patterns (e.g. transients) and spectral regularities (subcarrier occupancy, harmonic layouts) can be jointly exploited to estimate the signal of interest (SoI). In contrast to single-domain designs, the cross-domain attention layers allow the model to exchange information between temporal and spectral representations before decoding, improving robustness to semi-structured interference [12].

The backbone of the model is a **Bi-UNet model**. As illustrated in Fig 2, the *time branch* is a 1D encoder-decoder with strided convolutions for downsampling, transposed convolutions for upsampling, and skip connections at matched scales; it operates on two real-valued channels (I and Q). The *spectral branch* applies 2D convolutions over the STFT, reducing along the frequency axis until a single channel remains, then switches to time convolutions so both branches reach the same tensor shape at the bottleneck; the spectral output is later mapped back to time by iSTFT. Proper alignment of STFT window and temporal strides ensures the bottleneck tensors

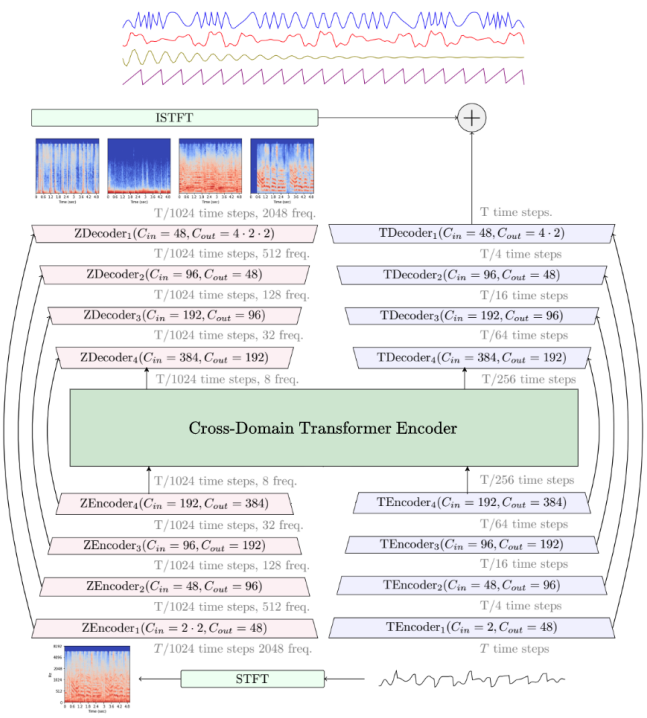


Fig. 2. HT-Demucs Architecture [12]

are adequate, enabling residual fusion and attention across domains.

The innermost layers of Hybrid Demucs are replaced by a Transformer encoder that interleaves self-attention within each domain and cross-attention across domains (the central green block in Fig. 2). Positional encodings (1D/2D), LayerNorm/LayerScale, multi-head attention and $4\times$ MLP expansion are used for stability and capacity. To extend the receptive field without quadratic memory, sparse attention kernels are employed, enabling long-context modeling for periodic structures in OFDM-like signals while keeping compute bounded.

HT-Demucs is trained end-to-end to predict the clean SoI $\hat{s}$ from the mixture y using time-domain losses (e.g., ℓ_1) optionally combined with multi-resolution STFT losses, so that both transient detail and spectral structure are matched. Training uses fixed-length segments; inference relies on overlap–add windowing (default overlap ≈ 0.25 in our setup) to control boundary artifacts and memory, consistent with Demucs practice.

For RF I/Q inputs, we represent real and imaginary parts as separate channels, choose STFT parameters consistent with the communication setting (e.g., FFT size and hop tied to OFDM subcarrier spacing and cyclic prefix), and let the spectral branch learn frequency-level structures while the time branch handles symbol-scale effects. Cross-domain attention then aligns temporal context with frequency-localized patterns.

1) UNet (review and link to HT-Demucs): Because UNet is the backbone of HT-Demucs, we briefly recap it in the 1D/2D setting. UNet is an encoder–decoder with symmetric skip connections that preserve fine-grained detail from early layers while the encoder aggregates context at coarser scales (Fig 3) [6] [7]. In 1D sequences it presented a fine multi-scale feature extraction with sample-level precision. HT-Demucs extended this idea with stronger backbones and hybrid time–frequency processing. It inherits these multi-scale/skip properties in both branches and augments the bottleneck with cross-domain attention, letting the model choose which domain (or combination) is most informative for a given interference scenario.

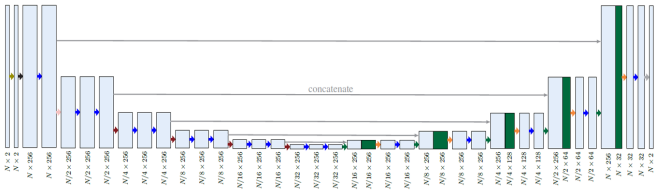


Fig. 3. UNet Architecture [6]

B. Isolated Branch Study: Time-Only and Spectral-Only Branches

To disentangle the contribution of each domain, we evaluate the two HT-Demucs branches in isolation: a temporal-only model (temporal UNet path) and a spectral-only model (STFT UNet path). This design choice is motivated by two factors. First, prior work and our own baselines confirm that UNet-type separators are strong in RF separation; HT-Demucs builds on that backbone but adds cross/self-attention and dual-domain fusion. Second, by reusing each HT-Demucs branch as a stand-alone network we can directly compare the individual branches versus the **hybrid** model, thereby quantifying what is gained by cross-domain coupling. Naturally, evaluating branches separately *removes* the cross-domain attention at the bottleneck, so any performance difference with respect to HT-Demucs can be ascribed to the absence of time–frequency fusion rather than to capacity or training changes.

1) Time-Only (Temporal Branch): The time-only model retains the 1D encoder-decoder used by the HT-Demucs temporal path: strided convolutions for downsampling, transposed convolutions for upsampling, and symmetric skip connections across scales, operating on the I/Q representation (real and imaginary parts as separate channels). This architecture is reflected in Fig 4. Compared to a plain UNet, this branch benefits from a deeper multi-scale down/upsampling schedule that yields a larger receptive field at the bottleneck and end-to-end time-domain optimization (paired with overlap–add inference) that preserves sample-level phase consistency. This branch measures how far a purely temporal separator can go on RF mixtures. In particular, it tests whether long-range temporal context suffices to recover a demodulable SoI across the different interference power values considered.

2) Spectral-Only (STFT Branch): The spectral-only model reuses the HT-Demucs frequency path: 2D convolutions over

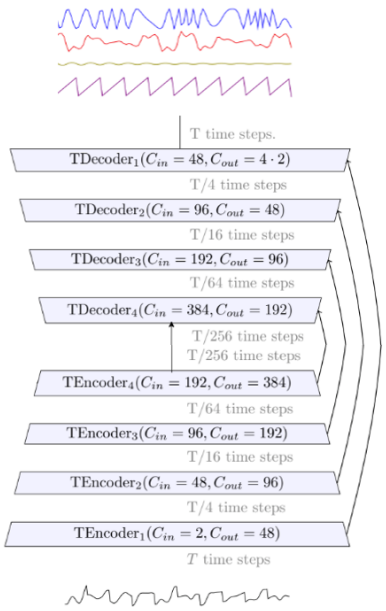


Fig. 4. HT-Demucs Temporal Branch Architecture [12]

complex STFTs that progressively contract the frequency axis until a single "frequency channel" remains, followed by time-conv layers, and an inverse STFT to return to the waveform domain at the output. In Fig 5 is shown. Training targets the complex spectrogram (or its real–imag parts) so that magnitude and phase information are explicitly modeled; the iSTFT ensures time-domain consistency at reconstruction [14]. This variant tests a different angle, since RF signals show local

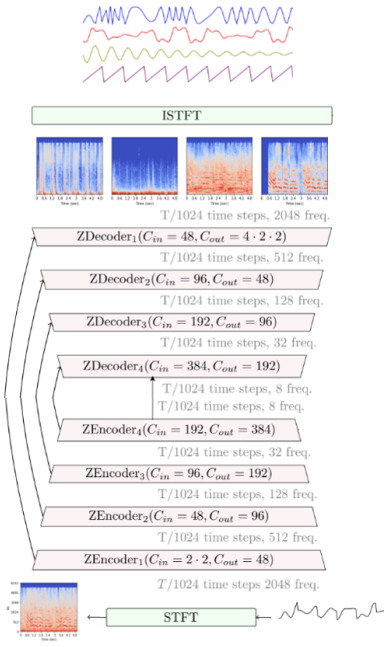


Fig. 5. HT-Demucs Spectral Branch Architecture [12]

patterns across subcarriers (active tones, pilots, notches) that the frequency view captures well. The question here is how well a purely spectral separator performs when deprived of temporal cross-attention: it should handle subcarrier-localized interference and multi-carrier structure, but may miss longer-term time relationships that the temporal branch learns better, and are, at first, essential in communications.

IV. Experimental Study

In this section, the experiments for single–channel RF signal separation across six mixture types that combine three multi–carrier SoIs with two Bluetooth interferences will be presented. We evaluate interference-rejection fidelity As in prior work, we evaluate both downstream decoding performance (bit error rate, BER) and interference–rejection fidelity (mean–squared error, MSE) of the denoised SoI. This section details the datasets and mixture–generation protocol used to ensure diversity and reproducibility.

A. Signal Types and Datasets

All datasets used in this work are generated with our internal framework [8], which builds synthetic but realistic baseband I/Q sequences directly from video transmissions. This ensures that every signal instance corresponds to a fixed information payload (one video clip), while still allowing flexibility in modulation type, waveform structure, and interference scenarios.

We consider three OFDM waveforms with different constellation orders:

- **OFDM–QPSK**
- **OFDM–64QAM**
- **OFDM–1024QAM**

For each signal of interest (SoI), we generate **100** sequences using **20** source videos and **5** non–overlapping clips per video (one clip per generated signal). Although each waveform carries the same amount of information (one video clip), the time duration varies with the modulation order, as higher–order constellations yield shorter sequences for the same payload.

As for the interferences, we use two types of Bluetooth technology:

- **Bluetooth BR** (Basic Rate)
- **Bluetooth LE** (Low Energy)

These interference sets are also generated through the same framework, again from video clips, to introduce realistic symbol dynamics. Their raw lengths are adapted to match each SoI instance at mix time. When inputs differ in sampling rate, we resample to a common rate (policy: choose the higher f_s to preserve bandwidth) and trim to a common time length before mixing. These steps are performed consistently by our dataset builder.

1) Test Dataset: To assess performance and cross-constellation generalization, the test set includes two SoIs and a single OFDM-based interferer:

- **SoIs: OFDM–64QAM** (seen during training) and **OFDM–256QAM** (unseen; high order).

- **Interference:** a **Wi-Fi** signal; OFDM-based waveform, chosen to introduce realistic, multi-carrier spectral structure.

The OFDM-64QAM SoI serves as a representative case from the modulations used in our training set (QPSK/64QAM/1024QAM), allowing a fair read on how well the separator performs when the constellation family matches training. In contrast, OFDM-256QAM was *not* used for training, so it tests generalization to new constellation orders within the same OFDM technology. Its higher order makes it more sensitive to residual distortion, making it a more challenging and informative target. The Wi-Fi interferer shares OFDM structure with the SoIs, which may increase separation difficulty without being a match to the SoI parameters.

As for the geneartion, the same internal framework as for training was used. For each SoI we generate **5** signals, each from a single clip from 5 different source videos (one clip per video). As before, each sample carries a fixed information payload (one video clip), so sequence duration is modulation-dependent. At mix time we resample to a common sampling rate when needed, trim to a common length, and construct mixtures with the Wi-Fi interferer following the same SINR targeting policy as in training. All signals are handled in baseband I/Q.

B. Mixture Construction and SINR Schedule

For each SoI/interferer pair we construct mixtures $y^{(i)}$ as

$$y^{(i)} \;=\; s^{(i)} \;+\; a_i\, b^{(i)} \;+\; n^{(i)}, \tag{2}$$

where a_i scales the interferer and $n^{(i)}$ is additive noise. We target a broad range of working points by sampling **SINR** levels uniformly from **0 to 30 dB** for training. We use the conventional definition [9]

$$\mathrm{SINR} \;=\; \frac{P_s}{P_i + P_n}, \tag{3}$$

and solve for the per–item parameters as follows:

1) Measure RMS powers $P_s = \|s\|_{\mathrm{rms}}^2$ and $P_i = \|b\|_{\mathrm{rms}}^2$ on the resampled, time–aligned frames.
2) Choose a_i and n_i (AWGN with variance σ^2, so that the desired SINR is met in power: $P_s/(a_i^2 P_i + \sigma^2) = 10^{\mathrm{SINR}/10}$; if the required $\sigma^2 \le 0$, we keep $\sigma^2{=}0$ (interference already dominates).

The complete procedure (resampling to a common f_s, trimming to a common length T_{out}, per–item SINR targeting, and reproducible logging of per–item parameters) is implemented in our mixture builder and recorded in a metadata file (*.json*)

We thus obtain six training/evaluation datasets:

- OFDM–QPSK + Bluetooth LE
- OFDM–QPSK + Bluetooth BR
- OFDM–64QAM + Bluetooth LE
- OFDM–64QAM + Bluetooth BR
- OFDM–1024QAM + Bluetooth LE
- OFDM–1024QAM + Bluetooth BR

All inputs are stored as baseband I/Q in HDF5 with consistent shapes and metadata, and are normalized at load time for stable training.

For the training, the model processes I/Q in both time and time–frequency domains with cross–domain attention (HT–Demucs). We organize experiments as separate runs (e.g., `qpsk_le`, `64qam_br`, ...), optionally warm–starting from the previous best checkpoint for efficiency.

As was introduced, 3 differnt variations of the HT-Demucs model will be tested: hybrid, temporal branch and spectral branch. For training, each of them uses a different loss:

- *Temporal Branch* is optimized with an ℓ_1 waveform loss. For a predicted SoI $\hat{\mathbf{s}}$ and target $\mathbf{s}$, the temporal loss is

$$\mathcal{L}_{\text{time}} = \|\hat{\mathbf{s}} - \mathbf{s}\|_1, \tag{4}$$

  computed element-wise over I/Q and time [15].
- *Spectral Branch* is optimized with a multi–resolution STFT (MR-STFT) magnitude loss. It uses STFT magnitudes at three resolutions $\mathcal{F} = \{256, 512, 1024\}$ with hop $n_{fft}/4$ and Hann window, applied after averaging I/Q to a mono trace (as implemented in our training loop) [16]:

$$\mathcal{L}_{\text{spec}} = \frac{1}{|\mathcal{F}|} \sum_{n \in \mathcal{F}} \big\| \,|\mathrm{STFT}_n(\hat{\mathbf{s}}_{\text{mono}})| - |\mathrm{STFT}_n(\mathbf{s}_{\text{mono}})|\, \big\|_1. \tag{5}$$

- *Hybrid model* averages both:

$$\mathcal{L} = \tfrac{1}{2}\,\mathcal{L}_{\text{time}} + \tfrac{1}{2}\,\mathcal{L}_{\text{spec}}. \tag{6}$$

It may attract attention as for why the spectral branch uses a different loss. This is because multi–resolution STFT magnitudes emphasizes the frequency–domain structure that matters most for RF separation and demodulation.

In addition, we use 3 complementary metrics:

a) Normalized MSE (NMSE): On the validation split we adopt a single, model-independent metric for fair selection and comparison across our three loss variants:

$$\mathrm{NMSE}_{\mathrm{dB}} = 10 \log_{10}\!\left(\frac{\|\hat{\mathbf{s}} - \mathbf{s}\|_2^2}{\|\mathbf{s}\|_2^2 + \varepsilon}\right). \tag{7}$$

NMSE measures the relative distortion of the separated waveform, is scale-invariant, and does not depend on the specific training loss, making it a stable criterion to monitor separation fidelity [17].

b) ΔSINR (SINR gain): On the test split we report the SINR gain to quantify how many decibels of interference + noise are effectively removed by the separator, providing an interpretable "before vs. after" view. For a mixture $y = \mathbf{s} + \mathbf{b} + \mathbf{n}$,

$$\mathrm{SINR}_{\text{in}} = \frac{\|\mathbf{s}\|_2^2}{\|y - \mathbf{s}\|_2^2},\ \mathrm{SINR}_{\text{in,dB}} = 10 \log_{10}(\mathrm{SINR}_{\text{in}}), \tag{8}$$

and after separation,

$$\mathrm{SINR}_{\text{out}} = \frac{\|\mathbf{s}\|_2^2}{\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2}, \quad \mathrm{SINR}_{\text{out,dB}} = 10 \log_{10}(\mathrm{SINR}_{\text{out}}). \tag{9}$$

We report

$$\Delta\mathrm{SINR}_{\mathrm{dB}} = \mathrm{SINR}_{\mathrm{out,dB}} - \mathrm{SINR}_{\mathrm{in,dB}}. \tag{10}$$

Therefore, a $\Delta\mathrm{SINR}_{\mathrm{dB}} = 0\,\mathrm{dB}$ means no improvement; positive values indicate suppression of interference/noise (e.g., $+20\,\mathrm{dB} \Rightarrow 100\times$ reduction in residual power), negative values indicate degradation [18].

c) Bit Error Rate (BER): We also evaluate end-to-end communication utility on the test split via uncoded BER. We demodulate the separated SoI back into bits and compare against the ground-truth payload used to generate each OFDM frame (video clips). BER is the fraction of incorrect bits; lower is better:

$$\mathrm{BER} = \frac{1}{N_b}\sum_{n=1}^{N_b} \mathbb{1}\left[b_n \neq \hat{b}_n\right] \tag{11}$$

Unlike waveform metrics (NMSE, ΔSINR), BER directly reflects whether the separation preserves information to the extent required for reliable decoding [1].

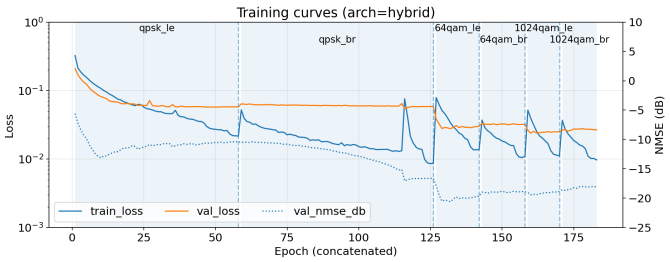
V. Results

Here we analyze performance in three stages. First, we inspect the training dynamics of the three separators (*temporal*, *spectral*, *hybrid*) by plotting *training loss*, *validation loss*, and the common validation $\mathrm{NMSE}_{\mathrm{dB}}$. It is shown in graphs from Fig 6. Second, we evaluate separation on the *test set* by sweeping the input SINR and reporting, for each model and for both datasets (OFDM–64QAM and OFDM–256QAM), the curves ΔSINR vs. SINR and $\mathrm{NMSE}_{\mathrm{dB}}$ vs. SINR, thereby capturing both relative interference suppression and absolute waveform fidelity. These results are shown in Fig 7. Finally, fixing $\mathrm{SINR} = 16$ dB, we perform end-to-end demodulation and compare BER across four conditions: the uncleaned mixture and the outputs of the hybrid, temporal, and spectral models (Fig 8). This graphics are complemented with representative visual frames (clean reference, interfered input, and cleaned outputs) to qualitatively corroborate the quantitative findings in Fig 9.

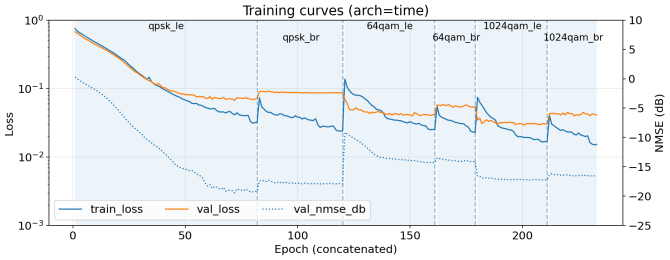
A. Results Analysis

Firstly, concerning the training across the six experiments, the *hybrid* separator (Fig 6a) converges the fastest and achieves the best waveform fidelity (lowest NMSE), with validation $\mathrm{NMSE}_{\mathrm{dB}}$ typically around -18 to -21 dB. The *temporal* model (Fig 6b) closely follows, stabilizing near -15 to -18 dB with small train-validation gaps, indicative of good generalization and limited overfitting. In contrast, the *spectral-only* (Fig 6c) variant reduces its spectral loss but has poor validation $\mathrm{NMSE}_{\mathrm{dB}}$ (about 0 to +5 dB), revealing a poor performance in I/Q waveform fidelity.

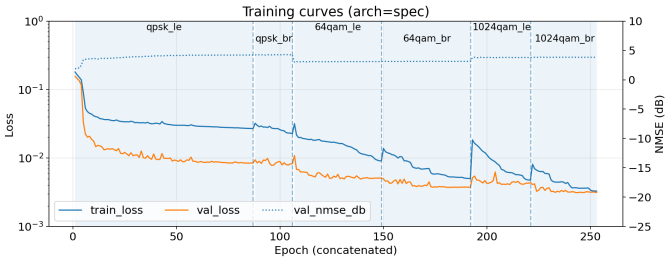
Concerning the $\Delta SINR$ and $NMSE$ vs. $SINR$ graphs (Fig 7), it can be highlighted how the *spectral* model has the largest $\Delta SINR$ value (suggesting a great interference suppression), while also having the largest $NMSE$ value (>2dB; indicating an important distorsion in the recovered waveform). The other models present a predictable behavior:



(a) Hybrid model



(b) Temporal Branch



(c) Spectral Branch

Fig. 6. Training curves for each model.

high and relatively constant $\Delta SINR$ (≈ 32dB) throughout all $SINR$ values and low descending-with-the-$SINR$ values for the $NMSE$, suggesting a good interference suppression.

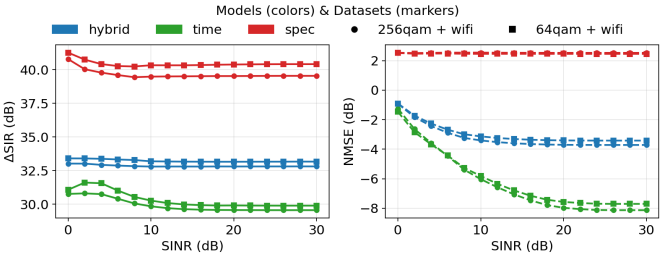
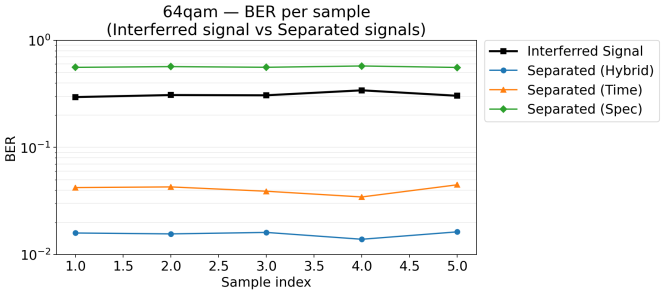


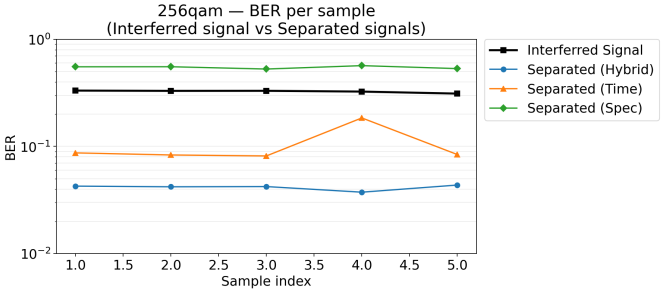
Fig. 7. NMSE and $\Delta SINR$ metrics for test dataset

And finally, when having demodulated the test signals and calculated their BER, from values depicted Fig 8, we can get a good intuition of the over-all results of the experiments: The one with best BER is the *hybrid* model, closely followed by the *temporal* one, while the *spectral* model presents worse BER than the initial interfered signal for both datasets. All this can be checked also by watching the demodulated frames

in Fig 9, where we see a pretty good recovery of the SoI for *hybrid* and *temporal* models. Also, we can actually see that the signal recovered for the *spectral* model has a BER of around 0.5, which indicates that the bitstream 'recovered' is completely random. On the OFDM-256QAM test set, our method markedly improves the recovered signal, although with smaller gains than for the seen OFDM-64QAM case. Even so, Fig.8b reports a BER of approximately 0.05, which is consistent with the nearly clean frame in Fig.9b and close to the quality achieved with OFDM-64QAM (Fig 9a), when looking at both images. This result is noteworthy given that 256-QAM's higher-order constellation is more sensitive to residual distortion and that the unprocessed mixtures start from a higher BER; despite these factors, the recovery remains strong.



(a) Test Dataset: OFDM-64QAM



(b) Test Dataset: OFDM-256QAM

Fig. 8. BER for test datasets.

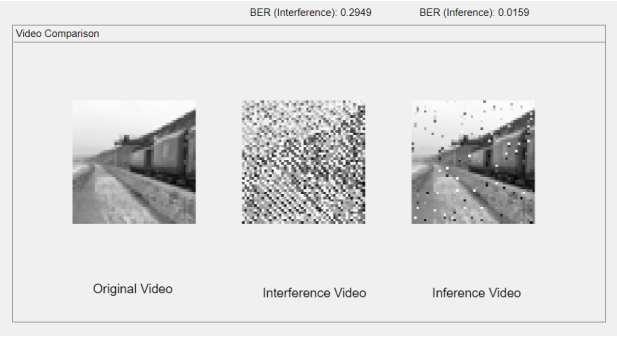
VI. Conclusions

This work studied data–driven interference mitigation for single–channel RF mixtures using a hybrid time–frequency separator adapted from HT–Demucs and two branch–isolated variants (time–only and spectral–only). Across controlled OFDM scenarios and realistic interferers, we evaluated separation fidelity (NMSE, ΔSINR) and communication utility (BER after demodulation), with qualitative checks on decoded video frames. The main conclusions we have reached are:

- **Hybrid time–frequency separation is the most reliable from the studied.** The hybrid model consistently achieves the best validation NMSE and the lowest BER on test signals, indicating that cross–domain attention effectively fuses complementary structure: symbol–scale temporal patterns in I/Q and subcarrier–localized regularities in the spectrogram.
- **Time–only is a strong, simple baseline.** Even though this was known from previous works (basically a UNet), we reaffirm that the temporal branch is reliable. This model alone closely tracks the hybrid model in NMSE and BER and delivers a large, relatively flat SINR gain across input SINR. For settings prioritizing simplicity and latency, the time path is a competitive choice.
- **Spectral–only shows a failure mode.** Despite often reporting the *largest* ΔSINR, the spectral model exhibits poor NMSE and high BER, frequently worse than the unprocessed mixture. This indicates that it suppresses broadband interference while introducing I/Q distortions catastrophic for demodulation.
- **Cross-correlation between branches works**. While the spectral-only mode fails to recover the interfered signal (and even worsens it), the *hybrid* model outperforms the *temporal-only* one. This improvement suggests that the cross-correlation between branches plays a key role in enhancing the model's performance.
- **Generalization to unseen constellations.** On OFDM–256QAM (unseen during training) the hybrid model (best performance) attains a BER of ≈ 0.05 (Fig. 8b) with near-clean reconstructions (Fig. 9b), substantially improving over the mixture despite a higher starting BER. The time-only branch follows the same pattern with smaller gains. This evidences meaningful modulation-order generalization, while confirming that higher-order constellations are more sensitive to errors, which narrows the margin relative to 64-QAM.
- **Dataset design enables robust evaluation.** The controlled dataset, built from video transmissions with support for multiple modulation formats and interference types, proved essential for assessing both in-distribution and out-of-distribution performance. Its flexibility allowed us to test generalization to unseen constellations (e.g., 256-QAM) under realistic conditions, ensuring that the observed behaviors are not artifacts of a narrowly tailored training set.

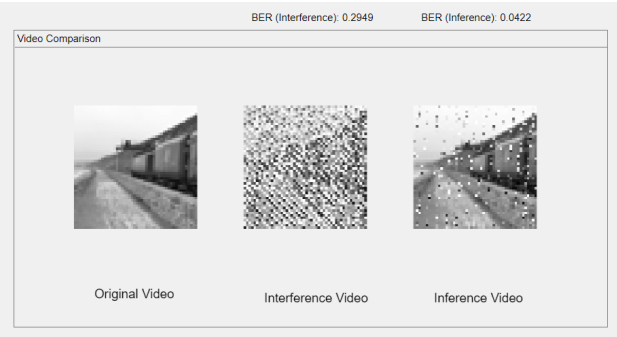
In practical terms, the separator could be deployed as a learned prefilter (hybrid or time-only) placed before a conventional demodulator. This drop-in design improves robustness in interference-limited regimes—particularly at moderate and low input SINR—without changing the downstream receiver chain, making it a realistic add-on for existing stacks.

Moreover, in many test cases the hybrid model yields a near-complete recovery of the SoI: the waveform remains demodulable and the decoded video is close to be indistinguishable from the clean reference. Viewed through the lens of semantic communications, this level of preservation is sufficient for end-to-end use: even if small waveform errors persist, the conveyed information is effectively intact and can be consumed directly.

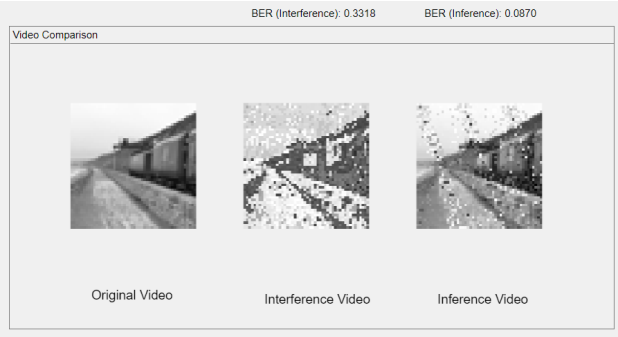


(a) OFDM-64QAM - Hybrid model

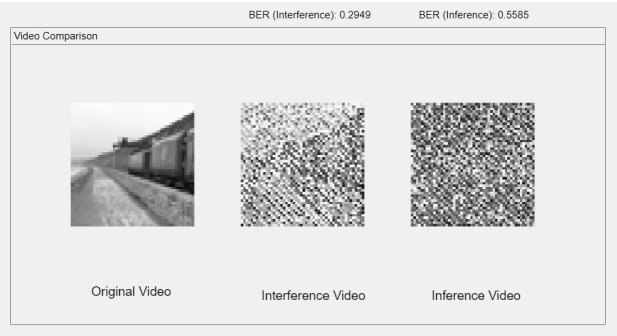
(b) OFDM-256QAM - Hybrid model



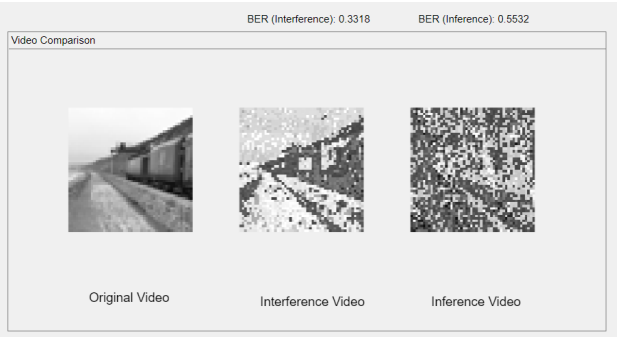
(c) OFDM-64QAM - Time model



(d) OFDM-256QAM - Time model



(e) OFDM-64QAM - Spectral model



(f) OFDM-256QAM - Spectral model

Fig. 9. Separated SoIs from interfered signal for each test dataset and for each trained model.

A. Future Works

In view of the achieved results for the stated problem, we could broaden our view and try to expand this project. Some future works might be:

- **Phase-aware learning:** Incorporate complex STFT or consistency losses, phase reconstruction objectives, and time–frequency consistency constraints to preserve demodulation-critical phase, which could help the spectral branch training, and therefore, maybe increase the performance of the hybrid model.
- **End-to-end & semantic objectives:** Jointly optimize separation with differentiable demodulation/decoding or BER proxies; add task-oriented objectives for semantic communications (for example: video intelligibility).
- **Broader robustness:** Extend to MIMO/spatial filtering, explicit channel impairments and online adaptation/domain shift.
- **Dataset-driven extensions:** Leverage the dataset framework to generate scenarios beyond OFDM+Bluetooth, e.g., 5G/6G waveforms, non-orthogonal multiple access, satellite links, or mixed-technology coexistence. Since the framework already supports several other modulation formats and interference types, these experiments can be systematically explored without redesigning the data pipeline.
- **Model unification & efficiency:** Train a single multi-interference, multi-modulation model; compress (distillation/quantization) for low-latency deployment on edge receivers.

VII. Data and Code Availability

All model code, including our HT-Demucs adaptations, training/inference scripts, and auxiliary utilities, is available in a public Git repository at https://github.com/mariogolbano/HTDEMUCS-for-interference-mitigation [19]. The dataset generation framework used to synthesize the OFDM/Bluetooth/Wi-Fi mixtures is hosted separately at https://github.com/mariogolbano/TFM-foundation-model-wireless-signals [8]. Both repositories include configuration files and step-by-step instructions to reproduce the experiments and figures reported in this work.

References

[1] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.

[2] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *arXiv preprint arXiv:2006.10685*, Jun 2020.

[3] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 363–378, Sep 1972.

[4] S. Tu, S. Chen, H. Zheng, and J. Wan, "Particle filtering based single-channel blind separation of co-frequency mpsk signals," in *Proc. 2007 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Xiamen, China, 2007, pp. 582–585.

[5] "Rf challenge: Data-driven signal separation in radio spectrum," MIT Research Laboratory of Electronics / USAF-MIT AI Accelerator, 2025, accessed: 2025-08-23. [Online]. Available: https://rfchallenge.mit.edu/

[6] A. Lancho, A. Weiss, G. C. F. Lee, T. Jayashankar, B. G. Kurien, Y. Polyanskiy, and G. W. Wornell, "Rf challenge: The data-driven radio frequency signal separation challenge," *arXiv preprint arXiv:2409.08839*, Jul 2025.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *arXiv preprint arXiv:1505.04597*, May 2015.

[8] M. Golbano, "Tfm-foundation-model-wireless-signals," GitHub repository, 2025, accessed: 2025-08-23. [Online]. Available: https://github.com/mariogolbano/TFM-foundation-model-wireless-signals

[9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press, 2005.

[10] N. T. Golmie, R. E. V. Dyck, and A. Soltanian, "Interference of bluetooth and ieee 802.11: Simulation modeling and performance evaluation," in *Proc. 4th ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Boston, MA, USA, 2001.

[11] A. F. Molisch, *Wireless Communications*, 3rd ed. Wiley–IEEE Press, 2022.

[12] S. Rouard, F. Massa, and A. Défossez, "Hybrid transformers for music source separation," *arXiv preprint arXiv:2211.08553*, Nov 2022.

[13] A. Défossez *et al.*, "facebookresearch/demucs: Code for the paper hybrid spectrogram and waveform source separation," GitHub repository, 2025, archived, accessed: 2025-08-23. [Online]. Available: https://github.com/facebookresearch/demucs

[14] D. Gabor, "Theory of communication," *Journal of the Institution of Electrical Engineers, Part III: Radio and Communication*, vol. 93, pp. 429–457, 1946.

[15] T. O. Hodson, "Root-mean-square error (rmse) or mean absolute error (mae): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, Jul 2022.

[16] R. Yamamoto, E. Song, and J.-M. Kim, "Parallel wavegan: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," *arXiv preprint arXiv:1910.11480*, Oct 2019.

[17] "Normalized mean squared error as a distance measure," Communications Toolbox Documentation, MathWorks, R2025a, 2025, accessed: 2025-08-23. [Online]. Available: https://www.mathworks.com/help/comm/ug/normalized-mean-square-distance-measure.html

[18] M. J. Rezaei, M. Abedi, and M. R. Mosavi, "New gps anti-jamming system based on multiple short-time fourier transform," *IET Radar, Sonar & Navigation*, vol. 10, no. 4, p. 807, Apr 2016.

[19] M. Golbano, "Htdemucs-for-interference-mitigation," GitHub repository, 2025, accessed: 2025-08-23. [Online]. Available: https://github.com/mariogolbano/HTDEMUCS-for-interference-mitigation