RF Challenge: The Data-Driven Radio Frequency Signal Separation Challenge

ALEJANDRO LANCHO [1,2] (Member, IEEE), AMIR WEISS [3] (Senior Member, IEEE),
GARY C. F. LEE [4] (Member, IEEE), TEJAS JAYASHANKAR[5] (Student Member, IEEE),
BINOY G. KURIEN [6], YURY POLYANSKIY [5] (Fellow, IEEE),
AND GREGORY W. WORNELL [5] (Fellow, IEEE)

[1]Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain
[2]Gregorio Marañón Health Research Institute, 28007 Madrid, Spain
[3]Bar-Ilan University, Ramat Gan 5290002, Israel
[4]Communications and Networks, Institute for Infocomm Research, Singapore 138632
[5]Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]MIT Lincoln Laboratory, Lexington, MA 02421, USA

CORRESPONDING AUTHOR: A. LANCHO (e-mail: alancho@ing.uc3m.es)

ABSTRACT We address the critical problem of interference rejection in radio-frequency (RF) signals using a data-driven approach that leverages deep-learning methods. A primary contribution of this paper is the introduction of the RF Challenge, which is a publicly available, diverse RF signal dataset for data-driven analyses of RF signal problems. Specifically, we adopt a simplified signal model for developing and analyzing interference rejection algorithms. For this signal model, we introduce a set of carefully chosen deep learning architectures, incorporating key domain-informed modifications alongside traditional benchmark solutions to establish baseline performance metrics for this intricate, ubiquitous problem. Through extensive simulations involving eight different signal mixture types, we demonstrate the superior performance (in some cases, by two orders of magnitude) of architectures such as UNet and WaveNet over traditional methods like matched filtering and linear minimum mean square error estimation. Our findings suggest that the data-driven approach can yield scalable solutions, in the sense that the same architectures may be similarly trained and deployed for different types of signals. Moreover, these findings further corroborate the promising potential of deep learning algorithms for enhancing communication systems, particularly via interference mitigation. This work also includes results from an open competition based on the RF Challenge, hosted at the 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'24).

INDEX TERMS Interference rejection, deep learning, source separation, wireless communication.

I. INTRODUCTION

THE RAPID proliferation of wireless technologies is driving an increasingly congested radio spectrum. Emerging services, such as virtual reality and augmented reality, demand substantial bandwidth to operate effectively [1]. Concurrently, new applications for ultra-reliable low-latency communications (URLLC) and massive machine-type communication (mMTC) are imposing

strict requirements on reliability, latency, and energy efficiency. These requirements necessitate advanced interference management strategies that go beyond traditional resource orthogonalization in time and frequency domains. As a result, wireless systems must adopt sophisticated interference management techniques to support these diverse, coexisting demands on the spectrum [2], [3].

Standard solutions for this ubiquitous problem involve filtering out interference by masking irrelevant parts of the time-spectrum grid or using multi-antenna capabilities to focus on specific spatial directions. In this paper, however, we focus on the case where the interference overlaps both in time and frequency with the signal of interest (SOI), and there is no spatial diversity to be exploited. Such a challenging case can occur, for example, in single-antenna devices or multi-antenna devices with insufficient spatial resolution to satisfactorily spatially filter the interference.[1] In such situations, judicious and effective solutions would have to exploit the specific underlying statistical structure of the interference, potentially via learning techniques.

Throughout this paper, we will also adopt the common terminology of *co-channel* interference to refer to other waveforms that operate at the same time and the same frequency band as the SOI [4]. Such co-channel interference can be reduced by the use of *interference mitigation* techniques, often via *signal separation* methods.[2] In this context, the goal is to extract the SOI with the highest possible fidelity, thereby enhancing downstream task performance (e.g., detection, demodulation, and decoding).

A. PREVIOUS WORK

The simplest solution for interference rejection in communication systems is to filter the received signal using a *matched filter* that is matched to the one used to generate the baseband signal waveform at the transmitter [5], thereby implicitly (and most likely incorrectly) treating the interference as additive white Gaussian noise (AWGN). Perhaps surprisingly, this is often the *only* interference mitigation method employed in existing wireless communication systems. However, it is well-known that the matched filter solution, while guaranteed to be optimal (in the maximum signal-to-noise ratio (SNR) sense) for an AWGN channel [5], is certainly not necessarily optimal in other settings. Consider, for example, an interference that is a communication signal generated from another communication system, overlapping with the SOI in time and frequency. In this case, in addition to (Gaussian) noise, the received signal will be contaminated with a non-Gaussian interference as well. In this scenario, matched filtering is likely to be suboptimal, thus creating the

possibility for other source separation techniques to provide performance gains.

There are, indeed, various source separation methods in the literature that were proposed and specifically designed for digital communication signals. One noteworthy approach is maximum likelihood sequence estimation of the target signal, for which algorithms such as particle filtering [6] and per-survivor processing algorithms [7] can be used. However, methods such as maximum likelihood, often referred to as "model-based" methods, require prior knowledge of the statistical models of the relevant signals, which may not be known or available in practical scenarios. As a result, in practice these methods are often suboptimal, and in some cases perform poorly (see, e.g., [8], [9]).

In those cases where the statistical models of the involved signals are not fully known, a more realistic (though challenging) paradigm is to assume that only a dataset of the underlying communication signals is available. This can be obtained, for example, through direct/background recordings or using high-fidelity simulators (e.g., [10]), allowing for a *data-driven* approach to source separation. In this setup, deep neural networks (DNNs) arise as a natural choice. This data-driven version of the source separation problem has been promoted within the context of the "RF Challenge" dataset, where signal separation with little to no prior information is pursued [11].

While machine learning (ML) techniques have shown promise in source separation within the vision and audio domains [12], [13], the radio-frequency (RF) domain presents unique challenges. Typically, these methods exploit domain-specific knowledge relating to the signals' characteristic structures. For example, color features and local dependencies are useful for separating natural images [14], whereas time-frequency spectrogram masking methods are commonly adopted for separating audio signals [15]. In contrast to natural signals, such as images or audio recordings, most RF signals are different in nature: i) they are *synthetically* generated via digital signal processing circuits; ii) they originate from discrete random variables;[3] and iii) they typically present an intricate combination of short and long temporal dependencies. On top of these differences, the mixture signals may overlap in time and frequency. All this together implies that classical, "handcrafted" model-based solutions—while successful in other domains—may fail in the RF signal domain (e.g., [17], [18]).

B. THE NEED FOR RF SIGNAL DATASETS

ML-based solutions often require large datasets. While large datasets are readily available in domains like vision and audio, they remain scarce in the RF domain, despite the widespread importance of digital RF communication signals in our everyday lives ([19, Ch. 2.4] and references therein).

---

[1]While we do not include technical details, it can be shown that the multi-antenna case can be effectively equivalent to a single-channel case after applying a beamforming vector to multivariate data from an array.

[2]We will henceforth refer to signal separation also as source separation or *interference rejection*, interchangeably. Furthermore, the interference rejection problem can also be understood as a denoising problem, where we aim to remove the SOI from the "noise", which, in this case, is the non-Gaussian interference signal.

[3]This is important because it implies that communication signals follow non-differentiable probability mass functions rather than differentiable probability density functions, which makes learning such distributions more challenging. These numerical challenges are further discussed in [16].

Among the available datasets, notable examples include the one provided by DeepSig, which offers several synthetically generated signals from GNU Radio for modulation detection and recognition [20], and those available through IQEngine, a Web-based software defined radio (SDR) toolkit for analyzing, processing, and sharing RF recordings [21].

A relatively new dataset designed specifically for source separation of RF signals is the "RF Challenge" [11]. This dataset includes several raw RF signals with minimal to no information about their generation processes. The lack of prior knowledge of the signal structure, combined with the possible complete overlap in time and frequency between the constituent signals, renders conventional separation via classical, and in particular linear, filtering techniques ineffective. Addressing this challenge calls for new learning methods and architectures [16], [22] that must go beyond the state of the art. The RF Challenge was created to promote the development of solutions to important problems particular to the RF domain, similar to how datasets such as MNIST, ImageNet, VAST, and HPC Challenge ([23], [24], [25], [26], respectively) have catalyzed research considerably in their respective areas by creating *standard benchmarks* and high-quality data.

The data associated with the RF Challenge are publicly available at https://rfchallenge.mit.edu/datasets/ and contain several datasets of RF signals recorded over the air or generated in lab environments. Specifically, most signals in the dataset are from the 2.4 GHz industrial, scientific, and medical (ISM) band. Only the 5G signals in the dataset were generated using a cable setup and a simulated 5G channel environment.

C. CONTRIBUTIONS

Our main contribution in this paper is the comprehensive presentation of the RF Challenge dataset for the *single-channel signal separation challenge*, focusing on two goals:

1) Separate a SOI from the interference;
2) Demodulate the (digital) SOI in such a mixture.

Rather than considering the classical formulation of source separation, we tackle this problem from a fresh, data-driven perspective. Specifically, we introduce a novel ML-aided approach to signal processing in communication systems, leveraging data-driven solutions empowered by recent advancements in deep learning techniques. These solutions are made feasible by progress in computational resources and the publicly available signal datasets we created and organized. We highlight that the methods developed within this research domain not only enable RF-aware ML devices and technology, but also hold the potential to enhance bandwidth utilization efficiency, facilitate spectrum sharing, improve performance in high-interference environments, and boost system robustness against adversarial attacks.

Through an extensive presentation of results, we show the potential of data-driven, deep learning-based solutions to significantly enhance interference rejection, and achieve improvements by orders of magnitude in both mean-squared error (MSE) and bit error rate (BER) compared to traditional signal processing methods. To support this claim, we introduce two deep learning architectures that we have established as benchmarks for interference mitigation, along with the performance results of the top teams from the "Data-Driven Radio Frequency Signal Separation Challenge" that we hosted at the ICASSP'24 Signal Processing Grand Challenges [27].

Finally, we conclude the paper by outlining a series of future directions focused on mitigating non-Gaussian interference in wireless communication systems. We expect these research directions, reinforced by competitions such as the recent SP Grand Challenge at ICASSP'24, to gain increasing relevance in the near future, and we invite researchers worldwide to actively contribute to advancing this field.

D. NOTATIONS

We use lowercase letters with standard font and sans-serif font, e.g., x and x, to denote deterministic and random scalars, respectively. Similarly, $\mathbf{x}$ and $\mathbf{x}$ represent deterministic and random vectors, and $\mathbf{X}$ and $\mathbf{X}$ denote deterministic and random matrices. Additionally, $\mathsf{x}[n]$ is used to represent the n-th random sample of the vector-form random signal $\mathbf{x}$. The uniform distribution over a set $\mathcal{S}$ is denoted as $\mathrm{Unif}(\mathcal{S})$, and for $K \in \mathbb{N}$, we define $\mathcal{S}_{1:K} \triangleq \{1, \ldots, K\}$. For brevity, we refer to the complex normal distribution as Gaussian. We denote $\mathbf{C}_{\mathsf{zw}} \triangleq \mathbb{E}[\mathbf{z}\mathbf{x}^{\mathsf{H}}] \in \mathbb{C}^{N_z \times N_w}$ as the cross-covariance matrix of the zero-mean vectors $\mathbf{z} \in \mathbb{C}^{N_z \times 1}$ and $\mathbf{w} \in \mathbb{C}^{N_w} \times 1$ (specializing to the auto-covariance $\mathbf{C}_{\mathsf{zz}}$ when $\mathbf{z} = \mathbf{w}$). The indicator function $\mathbb{1}_{\mathcal{E}}$ returns 1 when the event $\mathcal{E}$ occurs, and 0 event $\mathcal{E}$ occurs, and 0 otherwise.

II. PROBLEM STATEMENT

We consider the point-to-point, single-channel,[4] baseband signal model depicted in Fig. 1, where a transmitter aims to communicate a signal that carries a stream of encoded and modulated bits, referred to as the SOI and denoted as $\mathbf{s}$. The signal is measured at the intended receiver in the presence of an unknown interference signal, denoted as $\mathbf{b}$. The ultimate goal of the receiver is to successfully detect (or recover) the transmitted bits (or message) with the highest possible reliability, measured by the BER.

The input-output relation for a received, sampled, discrete-time baseband signal of length N samples is given by

$$\mathbf{y} = \mathbf{s} + \mathbf{b} \ \in \mathbb{C}^{N \times 1}. \tag{1}$$

This simplified model allows us to focus solely on the problem of interference rejection and the potential contributions of ML in this context. One can consider this model as the resulting input-output relation after successfully completing crucial processing stages in a communication

---

[4]The single-channel model encompasses scenarios such as single-antenna links and multi-antenna links where the spatial resolution is insufficient, resulting in a single effective channel between the transmitter and receiver.
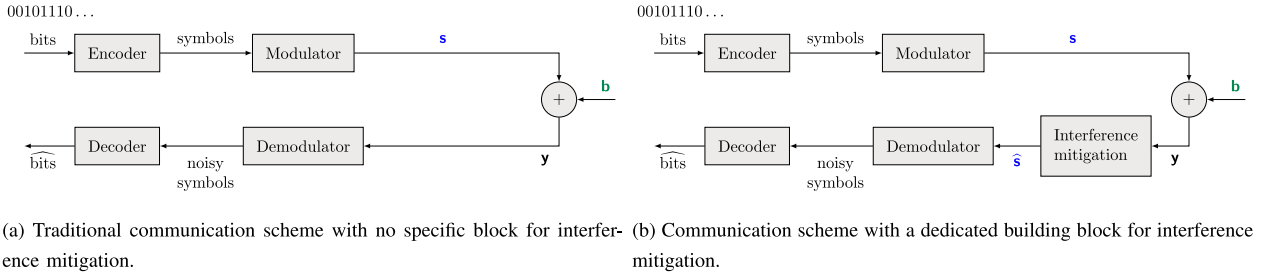
(a) Traditional communication scheme with no specific block for interference mitigation.

(b) Communication scheme with a dedicated building block for interference mitigation.

FIGURE 1. Communication schemes considered in this work.

system, such as time synchronization, channel estimation, and equalization. Although these aspects are deferred for future research, we acknowledge their importance in ensuring the correct operation of any practical communication system. Nonetheless, as we shall demonstrate throughout this paper, studying this building block in isolation enables us to understand the potential impact and challenges of integrating artificial intelligence (AI) capabilities into RF communication receivers.

Furthermore, we consider the case where the generation process of the interference signal **b** is unknown. Specifically, we assume that the interference consists of an unknown RF signal originating from another system operating in the same time-frequency band, possibly contaminated by AWGN.

Recall that we focus on digital communication signals as SOI in this work. In digital communication systems, the ultimate goal is to reliably recover the transmitted bits (or messages). Therefore, we consider the BER as a central measure of performance in this paper.

Note that we consider a scenario with non-Gaussian interference of unknown generation process, for which the optimal solution to minimize the BER is generally unknown.[5] Under this setting, various receiver architectural designs can be devised based on different principles, aiming to achieve the best possible BER performance. In this work, we propose a hybrid, "smart" receiver that first performs interference mitigation in a data-driven manner using a DNN. This approach aims to learn the relevant features of the unknown interference signal, as well as their statistical interactions with the relevant features of the SOI, in order to mitigate it. Then, by treating the residual interference as Gaussian noise, we apply standard matched filtering prior to decoding, so as to increase the postprocessing SNR. Consequently, we introduce a second measure of performance, namely the MSE between the estimated SOI $\hat{\mathbf{s}}$ and the true transmitted SOI **s**, to assess the signal quality after interference rejection and before decoding.[6]

---

[5]If the interference were Gaussian, applying a matched filter at the receiver, which is matched to the one used to modulate the encoded bits prior to decoding, would be optimal for the BER criterion (see Section III-A).

[6]Other performance measures could be considered depending on the specific receiver designs under consideration. Examples include packet-error rate when channel coding is part of the pipeline, peak-SNR to evaluate the fidelity of the reconstructed signal's amplitude, and outage probability, which is especially relevant for systems requiring guaranteed quality of service, such as URLLC.

A. SIGNAL MODELS

In this section, we categorize the various types of signals considered in this work based on our knowledge of their generation process.

When the signal's generation process is known, we have detailed information about the generated signal. Specifically, we can generate a synthetic dataset of signals for the sake of learning a data-driven source separation module. This approach is valuable when model-based solutions are infeasible, either because the model of the interference is unknown (but the SOI's model is known) or because the signal models are analytically intractable or such that lead to computationally impractical solutions. We will further categorize signals with a known generation process into single-carrier and multi-carrier signals.

When the signal generation process is unknown, we assume that we have datasets available, obtained through recordings or high-fidelity simulations. Thus, any knowledge relevant to performing source separation on these types of signals must be learned from the data.

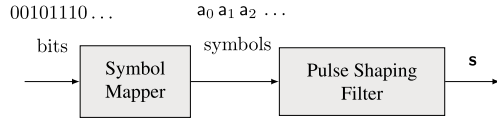### 1) SIGNALS WITH A KNOWN GENERATION PROCESS

We consider single-carrier and multi-carrier signals generated by linearly modulating symbols from constellations in the complex plane. Signals generated in this manner correspond to a prevalent class of digital communication signals observed in typical RF frequency bands.

Single-Carrier Signals: We consider single-carrier signals bearing M-bit long messages, which are mapped to L symbols from a given complex constellation (e.g., quadrature phase shift keying (QPSK)) using Gray coding. The bits are randomly generated via a fair coin toss and are all independent and identically distributed (i.i.d.). The n-th sample of $\mathbf{s} \in \mathbb{C}^{N \times 1}$ is expressed as

$$\mathsf{s}[n] = \sum_{\ell=0}^{L-1} \mathsf{a}_\ell \cdot g[n - \ell F - \tau_0], \qquad (2)$$

where $\mathsf{a}_\ell \in \mathcal{A}$ denotes a complex discrete symbol to be transmitted, with $\mathcal{A}$ being the constellation of (possibly complex-valued) symbols, $F \in \mathbb{N}$ is the symbol interval (in discrete-time), $\tau_0 \in \mathcal{S}_{0:F-1}$ is the offset for the first symbol, and $g[n]$ is the discrete-time impulse response of the transmitter filter (pulse shaping function). Figure 2

FIGURE 2. Block diagram for the generation process of the single-carrier signal, which modulates bits that are mapped into symbols from a complex-valued constellation before being filtered using a given pulse shaping filter.

FIGURE 3. Block diagram for the generation process of an OFDM symbol carrying symbols in each active subcarrier.

shows a simplified diagram for the generation process of the considered single-carrier signal type.

Multi-Carrier Signals: For multi-carrier signals, we focus on orthogonal frequency-division multiplexing (OFDM) signals, which are among the most widely used in key wireless communication technologies such as 5G and WiFi. An OFDM signal consists of K orthogonal sub-carriers, each carrying a symbol from a given (generally complex-valued) constellation [28]. We will consider the QPSK constellation for the numerical examples of this paper.

In this case as well, the bits are randomly generated using a fair coin toss in an i.i.d. manner and then mapped to symbols from the given constellation using Gray coding. The n-th sample of the SOI $\mathbf{s} \in \mathbb{C}^{N \times 1}$ is given by

$$\mathsf{s}[n] = \sum_{p=0}^{P-1} \sum_{k=0}^{K-1} \mathsf{a}_{k,p}\, r\big[n - p \cdot (K + T_{\mathrm{cp}}) - T_{\mathrm{cp}},\, k\big], \quad (3)$$

where

$$r[n,\, k] \triangleq \exp(j2\pi kn/K) \cdot \mathbb{1}_{\{-T_{\mathrm{cp}} \leq n < K\}}. \quad (4)$$
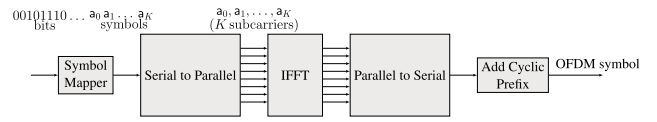
Here, K represents the total number of orthogonal complex exponential terms (subcarriers), where not all of them are necessarily active.[7] The value of K corresponds to the fast Fourier transform (FFT) size of the inverse FFT (IFFT) involved in generating an OFDM signal. The coefficients $\mathsf{a}_{k,p} \in \mathcal{A}$ are the information modulating symbols, where $\mathcal{A}$ represents the constellation.[8] A cyclic prefix (CP) is typically added before an OFDM symbol. Thus, each OFDM symbol is described within the interval $[-T_{\mathrm{cp}}, K]$, where T_{cp} is the CP length. The signals then span $P = N/(K + T_{\mathrm{cp}}) \in \mathbb{N}$ OFDM symbols, and their individual finite support is reflected by the finitely-supported function $r[n, k]$ in (4). Figure 3 illustrates the block diagram of the OFDM symbol generation process.

2) SIGNALS WITH UNKNOWN GENERATION PROCESS

When dealing with an unknown interference (e.g., from a different technology), accessing the signal generation process, which could potentially allow for the design of specific interference rejector, is often rare. However, one can rely on recorded interference signals to *learn* how to design

the interference rejector from the data. Another scenario where we may lack access to the generation process but still need to separate signals is the classical blind source separation problem. In this case, we may not know any of the signal models involved in the communication process, and our goal is simply to separate the superimposed signals into their constituent components. Finally, we could also consider the case where the generation process of the signals is known but is too complicated for deriving analytical solutions.

In any of these cases, the availability of signal datasets enables the design of data-driven source separators. We note that within the RF Challenge, there is a dataset of interference signals whose generative models are unknown, which can be used to develop learned, data-driven solutions.

III. METHODS

In this section, we review the methods used to perform interference rejection for the various combinations of SOIs and interference signals considered in this work. It is important to note that the signal models in this study are not necessarily known, making it impossible to derive theoretical performance bounds. Therefore, including a diverse set of numerically evaluated methods is essential. Alongside the proposed data-driven approaches, we include traditional methods that are widely used in both the literature and practical communication systems, providing well-established benchmarks for comparison.

A. TRADITIONAL METHODS

We now present two prevalent methods whose appeal comes from the balance between their theoretical justification—and, in fact, *optimality* for common criteria under certain conditions—and their simplicity, an important virtue in practical systems.

1) LINEAR MMSE ESTIMATION

A computationally attractive approach that exploits the joint second-order statistics of the mixture (1) and the SOI is optimal minimum mean-square error (MMSE) *linear* estimation. Assuming $\det(\mathbf{C}_{\mathsf{yy}}) \neq 0$ and that $\mathbf{s}$ and $\mathbf{b}$ are uncorrelated, the linear MMSE (LMMSE) estimator [5], given by

$$\widehat{\mathbf{s}}_{\mathrm{LMMSE}} \triangleq \mathbf{C}_{\mathsf{sy}}\mathbf{C}_{\mathsf{yy}}^{-1}\mathbf{y} = \mathbf{C}_{\mathsf{ss}}(\mathbf{C}_{\mathsf{ss}} + \mathbf{C}_{\mathsf{bb}})^{-1}\mathbf{y} \in \mathbb{C}^{N \times 1}, \quad (5)$$

is constructed using the second-order statistics of the mixture that inherently take into account the potentially nontrivial temporal structure of the interference expressed through $\mathbf{C}_{\mathsf{bb}}$.

---

[7]An "active" subcarrier is one that is being used to convey information, not necessarily random (e.g., pilots for the sake of channel estimation, and recall that pilots are predetermined, deterministic and known).

[8]For simplicity, the constellation includes the zero symbol, so (3) accounts for inactive subcarriers as well.

In other words, if $\mathbf{C_{bb}}$ somehow deviates from a scaled identity matrix, temporal cross-correlations exist.

While (5) coincides with the MMSE estimator when $\mathbf{y}$ and $\mathbf{b}$ are jointly Gaussian, it is generally suboptimal due to the linearity constraint. In our case, the signal $\mathbf{s}[n]$ is a digital communication signal and is certainly not Gaussian. As for $\mathbf{b}[n]$, its statistical model is assumed to be unknown throughout the design process of the interference mitigation module. Still, it would also typically be non-Gaussian, even if it contains AWGN, which is highly plausible.

Despite (5) not being the MMSE estimator in our scenarios of interest, it is still an important benchmark since it constitutes an attractive method for two main reasons. First, it is linear, and therefore fast and easy to implement for moderate values of N. Second, it *only* requires knowledge of second-order statistics, which are relatively easy to accurately estimate from data, even in real-time systems. We therefore use it as one of our benchmarks whenever it is computationally feasible.[9]

2) MATCHED FILTERING

Matched filtering, perhaps one of the most commonly used techniques in the signal processing chain of communication systems, exploits prior knowledge about the signal waveform (only) for enhanced detection of the transmitted symbols. When the residual (additive) component is Gaussian, it is optimal in the sense that it maximizes the SNR, and it is therefore also optimal in terms of minimum BER.

If the transmitted signal is represented by $\mathbf{s}[n] = \mathbf{a}_0 \cdot g[n]$, where $g[n]$ is the pulse shaping filter, then the matched filter would be $h_{\text{MF}}[n] = g^*[-n]$. In practical scenarios, the pulse $g[n]$ has finite duration. After performing the complex conjugation and time-reversal to obtain $g^*[-n]$, the resulting signal is shifted appropriately to ensure causality. This shift corresponds to aligning the start of the pulse with the beginning of the observation window.

This method is also an important benchmark as it is probably still the most commonly used method for symbol detection, which is the natural choice when the residual component, be it noise or interference, is treated as AWGN.

To conclude this section, we note in passing that the (theoretically naïve) option of not applying an interference mitigation method, namely only applying a matched filter to the received signal (1), will also be considered in our simulation as a benchmark. Indeed, with the complete absence of prior knowledge of the statistical model of the interference, this plain option of simply ignoring the interference may, after all, be chosen for practical considerations. While we do not advocate for such a solution approach, we acknowledge it as a realistic (even if not a leading) benchmark.

---

[9]For nonstationary input signals, the required inversion of $\mathbf{C_{yy}}$ is computationally impractical at high dimensions—matrix inversion (without a particular structure to be exploited) is generally of complexity $\mathcal{O}(N^3)$.

B. DATA-DRIVEN METHODS

This section presents the two most effective architectures we identified for data-driven source separation of RF signals: UNet and WaveNet. This selection was informed by insights from prior work. Specifically, in [22] we analyzed cyclostationary Gaussian signals to in order to isolate the effect of temporal correlations (i.e., second-order statistics) from higher-order statistics, and determine key modifications to standard deep learning architectures. This enabled us to compute the (exact) optimal MMSE solution, which was computable in this setup. We observed that extending the kernel size in the initial layer of the UNet—proportional to the effective length of the cross-correlation between the signals—significantly improved performance, closely approximating the optimal MMSE estimator in the cyclostationary Gaussian signal scenario. These findings guided our modifications of the UNet architecture, enhancing its performance when applied to real-world RF signals from the RF Challenge dataset.

Additionally, in [17] we evaluated state-of-the-art architectures from the audio domain on scenarios involving superimposed OFDM signals, now addressing separation based on higher-order statistics, where theoretically perfect separation was possible. However, these architectures struggled without domain-specific modifications. To address this, we introduced structural changes that led to up to a 30 dB improvement in separation performance. These modifications involved extending the kernel size in UNet-like architectures, so as to align with the OFDM's underlying FFT size, and employing dilated convolutions, which also motivated the inclusion of WaveNet in our subsequent works [16], [27].

Throughout the course of our research, we evaluated additional architectures, many of which did not yield notable improvements. Comparative performance results for these models are available in a dedicated GitHub repository for further reference.[10]

We henceforth assume we have a dataset of D i.i.d. copies of $\{(\mathbf{y}^{(i)}, \mathbf{s}^{(i)})\}_{i=1}^{D}$, i.e., the baseband versions of the mixture and SOI, whose real and imaginary parts are their in-phase and quadrature components, respectively.

1) UNET

The UNet, as depicted in Fig. 4, is a type of DNN originally proposed for biomedical image segmentation [29]. Its versatility has led to its adoption in various other applications, including spectrogram-based RF interference cancellation [30] and audio source separation [12], [31]. These applications typically correspond to a multivariate regression setup with identical dimensions for both input and output data.

Similarly to these aforementioned works, our approach employs 1D-convolutional layers to better capture the temporal features of time-series data. To effectively handle (baseband) complex-valued signals, inspired by widely

---

[10]https://github.com/RFChallenge/SCSS_DNN_Comparison

FIGURE 4. The UNet DNN architecture proposed for single-channel source separation of communication signals. The parameter κ denotes the kernel size of the first layer.

linear estimation techniques [32], we represent the real and imaginary parts as separate input channels. The UNet architecture comprises downsampling blocks, which operate on progressively coarser timescales, and incorporates skip connections to combine features from different timescales with the upsampling blocks.

It is well known that the careful design of a neural network architecture, tailored to the specific application, can significantly impact performance, as demonstrated by our experiments and architectural choices. Specifically, unlike standard CNN-based architectures tailored for image processing, which originally employed short kernels of size 3 in all layers, our UNet architecture features a first convolutional layer with a nonstandard, comparatively long kernel (indicated by κ in Fig. 4), which is of size 101—a difference of two orders of magnitude. We observed that proper adjustment of this hyperparameter to capture the effective correlation length of both the SOI and interference facilitates (and perhaps enables, as some of our findings indicates) the extraction of additional long-scale temporal structures of both signals, leading to performance gains of an order of magnitude compared to the originally proposed UNet [22].[11]

2) WAVENET

The WaveNet architecture [33] was initially introduced as a generative neural network for synthesizing raw audio waveforms. In subsequent work, it was adapted for the task of speech denoising [34]. At its core, the architecture uses stacked layers of convolutions with gated activation units. Unlike the downsampling and upsampling networks used in UNets, WaveNet preserves the temporal resolution at each layer while expanding the temporal receptive field by using dilated convolutions. As shown in Fig. 6, a dilated convolution can be interpreted as a kernel with spacing between elements, allowing the model to capture longer temporal dependencies without downsampling the sequence. For example, a dilated convolution with a kernel width of 3 and a dilation of 2 has an effective receptive field of 5.

---

[11] Our proposed UNet architecture for source separation of RF signals can be found at https://github.com/RFChallenge/icassp2024rfchallenge/blob/0.2.0/src/unet_model.py.

FIGURE 5. The WaveNet DNN architecture proposed for single-channel source separation of communication signals.

FIGURE 6. A dilated convolution operation with a kernel width of 3 and a dilation rate of 2, which results in a receptive field of 5.

As illustrated in Fig. 5, the WaveNet employs R residual blocks with dilated convolutions, where the output of block $i-1$ serves as the input to block i, for $i \in \{0, \ldots, R-1\}$. The dilated convolutions assist in learning long-range temporal and periodic structures. The dilations start small and successively increase, such that the dilation at block i is given by $2^{i \bmod m}$, where m is the dilation cycle length. For example, if the dilation periodicity is $m = 10$, then in block $i = 9$ the dilation is 512, and in block 10 the dilation is reset to 1. This allows the network to efficiently trade off between learning local and global temporal structures. All residual blocks use the same number of channels, C. Our WaveNet specifically uses $R = 30$ residual blocks, with a dilation cycle $m = 10$, and a number of channels per residual block of $C = 128$.

A few key modifications were made to facilitate training with RF signals compared to the original WaveNet [33]. First, since we are dealing with complex-valued continuous waveforms, we train on two-channel signals where the real and imaginary components of the RF signals are concatenated in the channel dimension. Second, we train with an MSE (squared ℓ_2) loss, as we did with the UNet. We monitor the validation MSE loss, and once the loss stops decreasing substantially, we stop training early. Lastly, we increased the channel dimension up to $C = 128$ to learn complex RF signals such as OFDM signals. Additionally, during data loading, we perform random time shifts and phase rotations on the interference to gain diversity and simulate typical transmission impairments in RF systems.[12]

IV. RESULTS

We present a diverse set of results for RF signal separation, examining various mixtures of signal types. Each combination exhibits unique joint statistical properties, introducing different levels of complexity in the task of learning effective signal separators.

We compare the performance of several approaches, including data-driven, neural network-based separators, as well as more traditional, commonly used methods. Beyond showcasing our contributions in developing ML-enhanced RF signal separation architectures, these results are also crucial for establishing standardized benchmarks that will serve as baselines for future research in this emerging field.

To analyze decoding capabilities (in terms of BER) alongside interference rejection capabilities (in terms of the MSE of the "denoised" SOI), we consider SOIs with known generative processes in this work. Specifically, we consider two different SOIs and four types of interferences, resulting in eight different combinations of mixture types, each of length $N = 40,960$ samples.

For the SOIs, we have:

1) *QPSK:* A single-carrier QPSK signal with an oversampling factor of $F = 16$, modulated by a root-raised cosine pulse shaping function with a roll-off factor of 0.5 that spans 128 samples (8 QPSK symbols due to the employed oversampling factor). We further apply an offset for the first symbol of $\tau_0 = 8$ samples. See Fig. 2 for a simplified diagram of the generation process of this SOI, which we refer to as "QPSK".

2) *OFDM-QPSK:* An OFDM signal where each subcarrier bears a QPSK symbol. We refer to this signal as "OFDM-QPSK". We set $T_{\text{cp}} = 16$, $K = 64$ subcarriers, with 56 active subcarriers (i.e., the 8 inactive subcarriers "carry" the zero symbol). Recall that these quantities were defined in (3). A simplified diagram of the generation process of this SOI is shown in Fig. 3.

The following four types of interference signals are only available through provided recordings, hence their generation process is unknown:
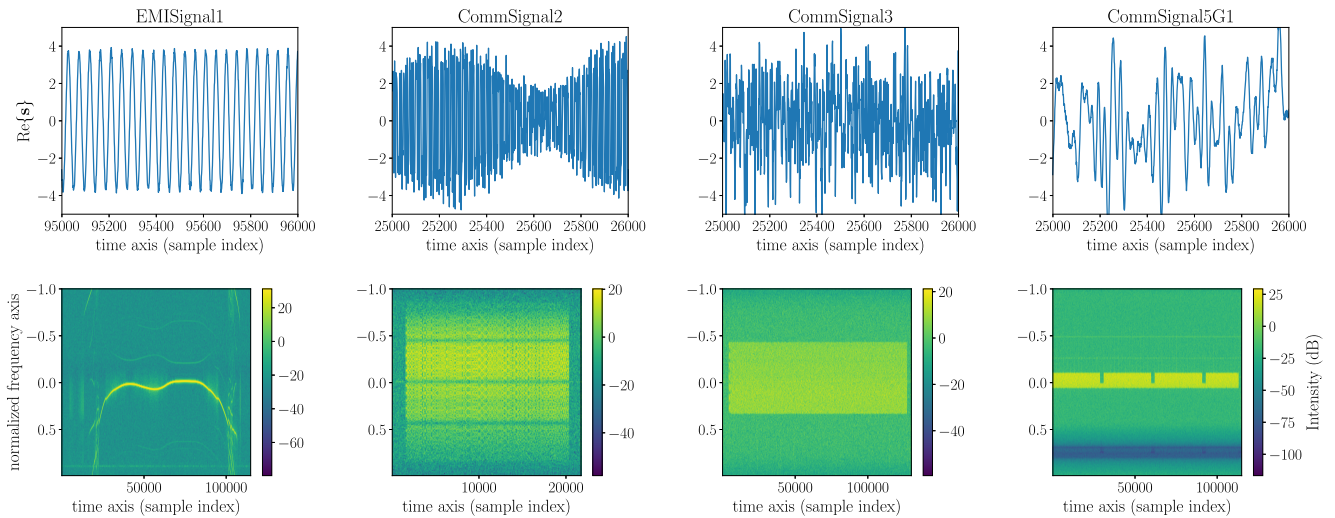
1) *EMISignal1:* Electromagnetic interference from unintentional radiation from an unknown RF-emitting device with a recording bandwidth of 25 MHz.

2) *CommSignal2:* A digital communication signal from a commercially available wireless device with a recording bandwidth of 25 MHz.

3) *CommSignal3:* Another digital communication signal from a commercially available wireless device with a recording bandwidth of 25 MHz.

4) *CommSignal5G1:* A 5G-compliant waveform with a recording bandwidth of 61.44 MHz.

We emphasize that the generative processes of the signals above are not only considered unknown in the simulations, but are in fact truly unknown to the authors. The dataset examples for the first three types (EMISignal1, CommSignal2, and CommSignal3) were recorded over-the-air, while the last one (CommSignal5G1) was generated and recorded within a controlled wired laboratory environment, with wireless impairments introduced via simulators.

To create interference signal examples, we divided the set of examples into training and test sets. A frame of the respective interference type was then randomly selected (uniformly) from the corresponding set, and a random window of $N = 40,960$ samples was extracted. Each interference component was scaled to achieve a target (empirical) signal-to-interference-and-noise ratio (SINR). Since all signal datasets are normalized to have unit power, for a target SINR level $\kappa^2 = 10^{(\text{SINR in dB})/10}$, the interference signal is scaled by $1/\kappa$. Each interference frame $\mathbf{b}^{(i)}$ also undergoes a random phase rotation before being added to the SOI $\mathbf{s}^{(i)}$ to create a mixture example $\mathbf{y}^{(i)}$ (see (1)). Note that we choose the term SINR rather than SIR since some of the signals we use were recorded, thus they inevitably contain additive noise. Consequently, the effective interference in these cases is a sum of a non-Gaussian interference component and an additive noise component.

For the recorded interference signals, the number of examples available per signal type changes. Furthermore, while the length of each recorded frame is the same for each signal type, it varies across types as well. In particular, we have 530 examples of 230,000 samples for EMISignal1, 100 examples of 43,560 for CommSignal2, 139 examples of 260,000 for CommSignal3, and 149 examples of 230,000 for CommSignal5G1. For consistency, we set the length of all input mixtures to 40,960 samples. Note that our dataset consists of signals represented in baseband form, where the carrier frequency component was removed prior to saving. However, since the signals were not precisely centered around the presumed carrier frequency, their spectral content was not perfectly aligned around zero. This resulted in a non-zero DC level and only partial frequency overlap. To simulate the intended setting of maximal frequency overlap,

---

[12]Our proposed WaveNet architecture for source separation of RF signals can be found at https://github.com/RFChallenge/icassp2024rfchallenge/blob/0.2.0/src/torchwavenet.py.

FIGURE 7. Representative frames of the four interference signal types in the dataset: EMISignal1, CommSignal2, CommSignal3, and CommSignal5G1. Top: Real part of the waveforms, $\mathrm{Re}\{s\}$; Bottom: Spectrogram of the respective signal frames.

we applied an additional frequency-shifting step to align the signals' spectral content around zero, ensuring their proper alignment for the analysis presented in this work. In particular, signals EMISignal1 and CommSignal5G1 were shifted in frequency to have their spectral energy content lie in baseband frequencies, simulating co-channel interference that overlaps both in time and frequency. Figure 7 shows the time- and frequency-domain representations of the recorded signal datasets used as interferences. Code examples can be found at https://rfchallenge.mit.edu/icassp24-single-channel/.

A. OUR RESULTS

Figures 8–11 show the performance of the two traditional interference rejection algorithms, introduced in Section III-A, and our proposed deep learning-based interference rejection algorithms, introduced in Section III-B, over the eight possible SOI-interference combinations. The performance is measured in terms of BER or MSE as a function of the target SINR. The plots include the following curves:

- *MF:* ignore the (potential) non-Gaussianity of the interference, and apply a matched filter to the mixture (1). We note that here matched filtering (MF) is applied only for the purpose of decoding (BER plots only).

- *LMMSE:* the SOI is estimated via MSE-optimal linear estimation, the best-performing traditional method described in Section III-A. Since the LMMSE requires the inversion of the covariance matrix $\mathbf{C_{yy}} \in \mathbb{C}^{N \times N}$, which is generally of complexity $\mathcal{O}(N^3)$, we have implemented this solution by applying (5) to consecutive blocks of length $2,560$ samples each.[13]

[13]Computing the LMMSE for sequences of length $40,960$ samples is impractical, as it requires inverting a $40,960 \times 40,960$ matrix, leading to computations on the order of $40,960^3 \sim 10^{12}$. However, we computed the LMMSE using blocks of length $2,560$, which is already close to its asymptotic value and, in particular, to the LMMSE for sequences of length $40,960$ samples.

- *UNet and WaveNet:* our proposed architectures, as presented in Section III-B. We emphasize that a separate neural network was trained for each mixture case.

Both the LMMSE estimation approach and DNN-based interference rejection methods include a final step that treats residual interference as Gaussian, applying standard matched filtering prior to decoding to improve post-processing SNR.
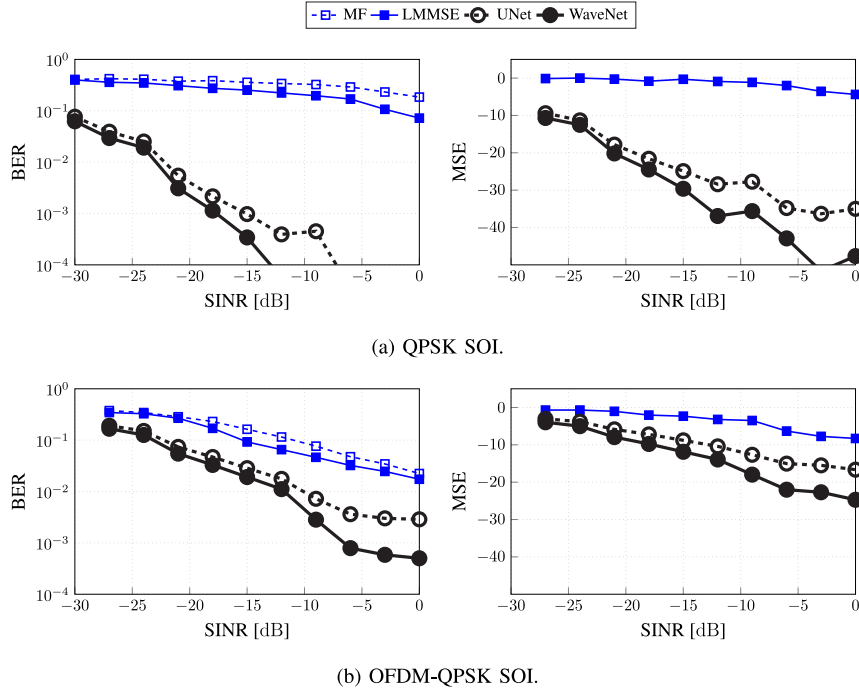
For the QPSK SOI case, this involves matched filtering, sampling at optimal points (assuming all necessary synchronization information is available), and then hard decoding to obtain symbols, which are mapped to bits. Similarly, for the OFDM SOI, we assume perfect synchronization, remove the cyclic prefix, apply an FFT of the appropriate size, and estimate the received symbols on active subcarriers, mapping them to their corresponding bits.

Both learning-based solutions described in Section III-B outperform the best traditional method out of the ones we consider as our benchmarks, namely LMMSE estimation, achieving up to two orders of magnitude performance improvements at considerably low SINR values. For instance, in Fig. 11(a), at an SINR level of $-18\,\mathrm{dB}$, the WaveNet model achieves a BER of approximately 10^{-3} and an MSE below $-20\,\mathrm{dB}$. In contrast, the solutions based on LMMSE and no mitigation (other than MF) only reach a BER slightly above 10^{-1} and an MSE around $0\,\mathrm{dB}$.
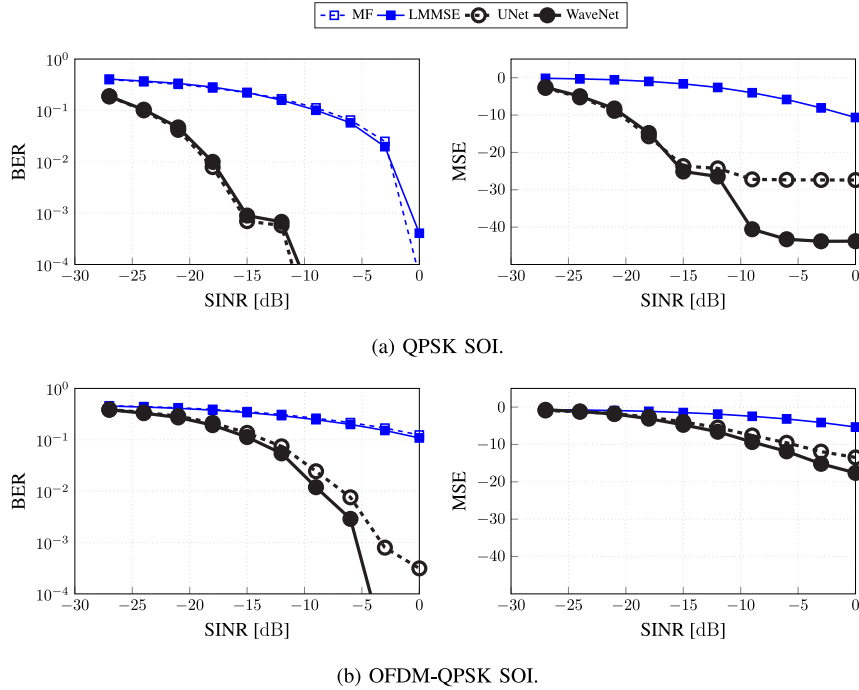
B. ICASSP'24 SP GRAND CHALLENGE RESULTS

We hosted the "Data-Driven Radio Frequency Signal Separation Challenge" as part of the ICASSP'24 Signal Processing Grand Challenges [27]. Among the submissions, only a few significantly outperformed the learning-based benchmark methods described in Section III-B at specific mixture cases [35], [36]. The proposed solutions are summarized below:

- KU-TII [35]: This team enhanced the WaveNet architecture provided as a benchmark. Their main contributions
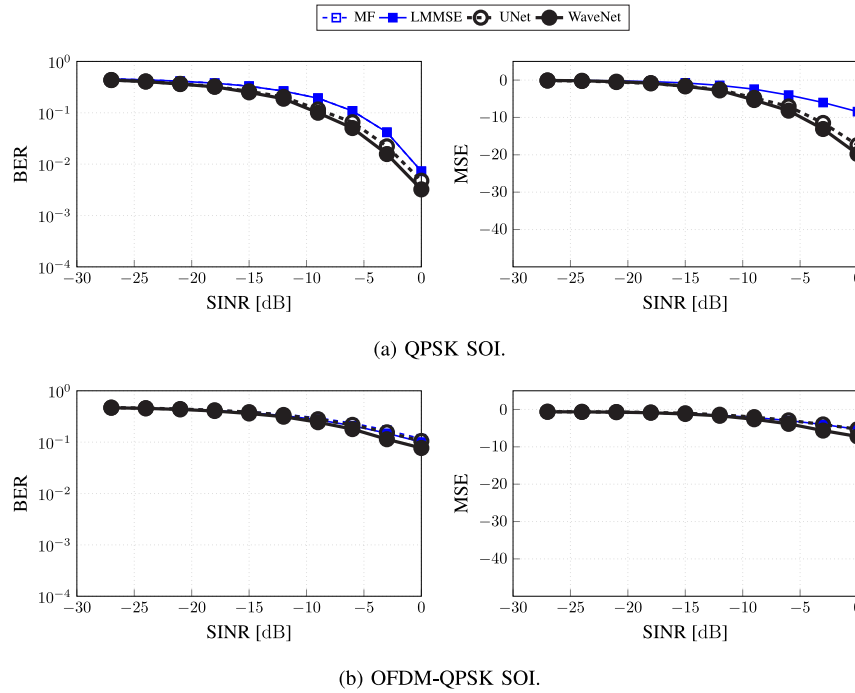
(a) QPSK SOI.

(b) OFDM-QPSK SOI.

FIGURE 8. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with EMISignal1 interference. This figure shows the benchmarks discussed in this paper, along with our proposed data-driven methods.
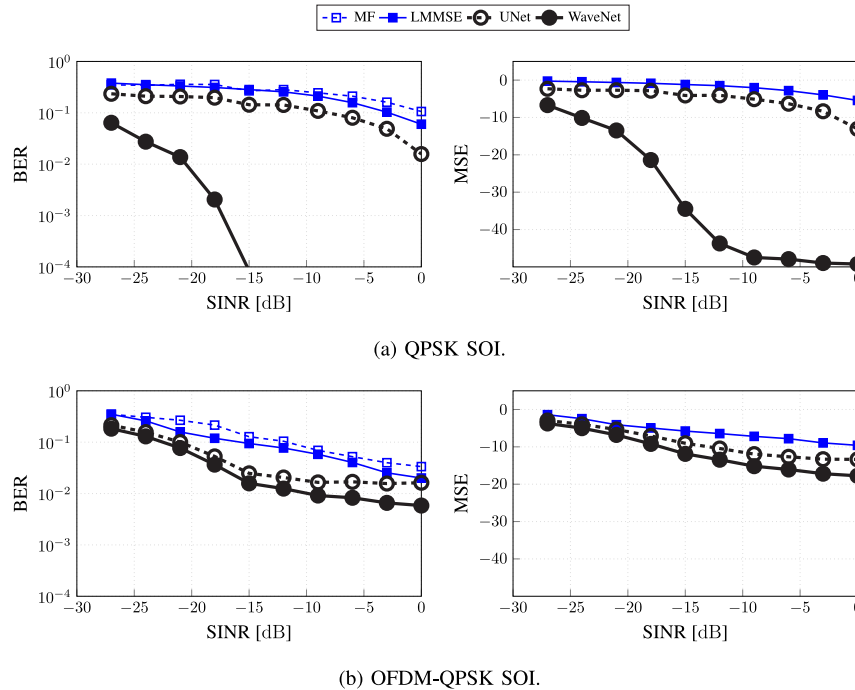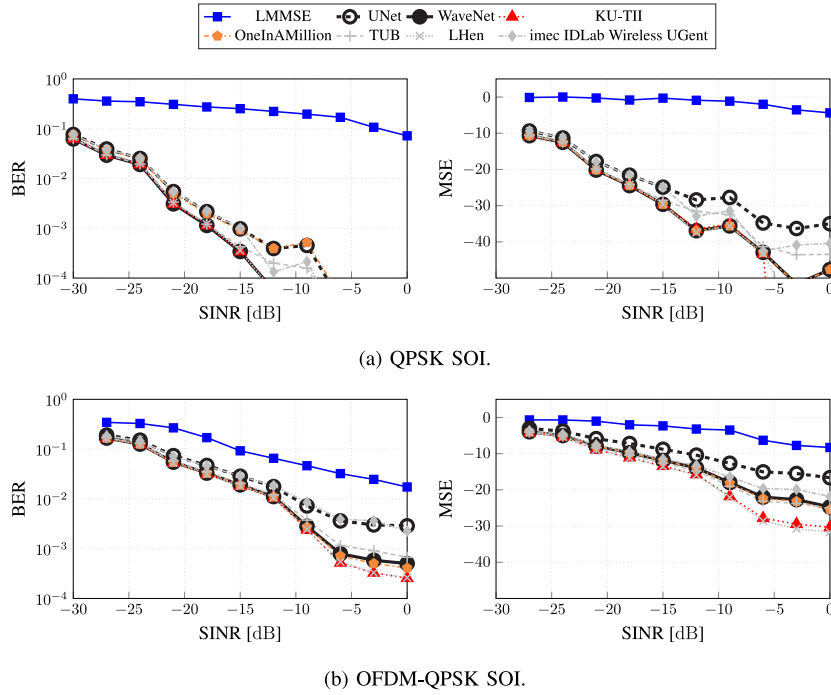

(a) QPSK SOI.

(b) OFDM-QPSK SOI.

FIGURE 9. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with CommSignal2 interference. This figure shows the benchmarks discussed in this paper, along with our proposed data-driven methods.

were two-fold: (i) Model-related: introducing learnable dilated convolutions within the convolutional layer; (ii) Data augmentation: expanding the training set by using as a validation set the test example data provided by the challenge organizers, and generating additional training examples for CommSignal2 by converting high-SNR waveforms with probably zero BER into bits, then reconverting these bits back into waveforms to subtract them from the mixtures. With this data augmentation strategy, the team trained their algorithm with more

(a) QPSK SOI.

(b) OFDM-QPSK SOI.

FIGURE 10. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with CommSignal3 interference. This figure shows the benchmarks discussed in this paper, along with our proposed data-driven methods.

(a) QPSK SOI.

(b) OFDM-QPSK SOI.

FIGURE 11. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with CommSignal5G1 interference. This figure shows the benchmarks discussed in this paper, along with our proposed data-driven methods.

data, potentially resulting in further performance gains unrelated to architectural improvements.

- One-In-A-Million [36]: This team presented two approaches, a Transformer UNet and a finetuned discriminative WaveNet. The Transformer UNet is a

convolution-attention-based model with an encoder-decoder architecture that includes self-attention blocks in the bottleneck to refine representations, similar to the one introduced in [37]. Between these two options, the finetuned WaveNet achieved superior

FIGURE 12. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with EMISignal1 interference. This figure includes the LMMSE benchmark, along with our proposed data-driven methods and the proposed methods of the participants of the ICASSP'24 SP Grand Challenge.

performance across all signal mixture cases, suggesting that Transformer architectures may require specific design modifications to fully leverage their capabilities in the source separation of digital communication signals.

- LHen [38]: This team extended the WaveNet baseline by incorporating an autoencoder tailored to the modulation type of the SOI. The encoder learns to demodulate the waveform estimated by WaveNet, while the corresponding decoder re-synthesizes the SOI waveform from the extracted bit sequence to achieve low MSE.
- TUB [39]: This team employed a DEMUCS architecture, featuring an encoder-decoder framework with a convolutional encoder, bidirectional LSTM applied on the encoder output, and a convolutional decoder, all linked via UNet skip connections [40], [41]. They further adapted DEMUCS for direct bit regression, similar to the "Bit Regression" baseline from the "Single-Channel RF Challenge" hosted on the RF Challenge website in 2021.[14]
- imec-IDLab [42]: This team developed a UNet architecture, building upon the baseline UNet provided in the challenge, redesigned specifically for separating interference signals in the time-frequency domain. For example, they leveraged domain-specific knowledge of SOIs for the OFDM-QPSK SOI case by integrating elements of OFDM signal resource grid configurations,
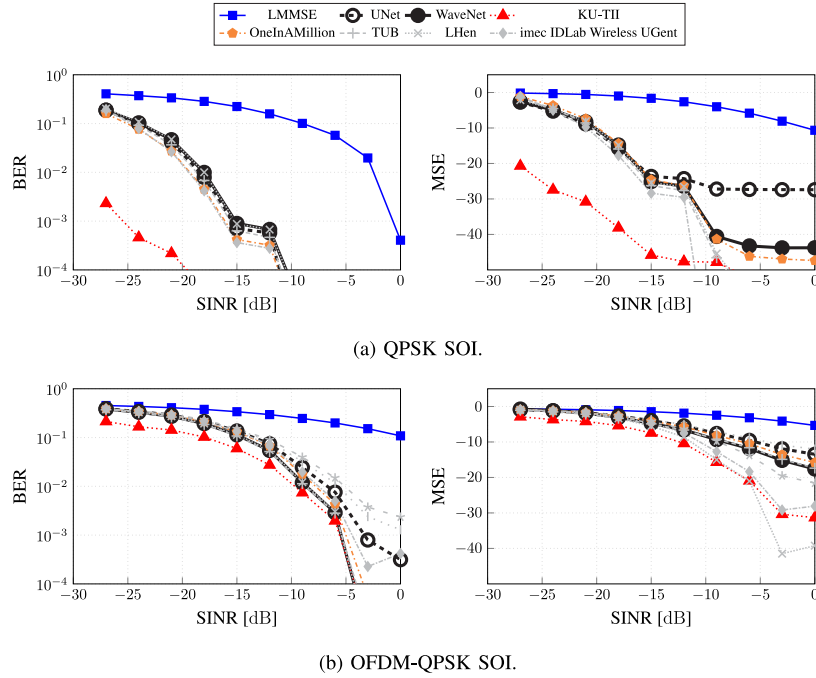
such as the cyclic prefix, into the architecture responsible for the decoding process.

Similar to the previous section, Figures 12–15 show the performance of the traditional interference rejection algorithm based on LMMSE estimation, our proposed deep learning-based interference rejection algorithms introduced in Section III-B, and the top-5-performing teams that participated in the challenge.
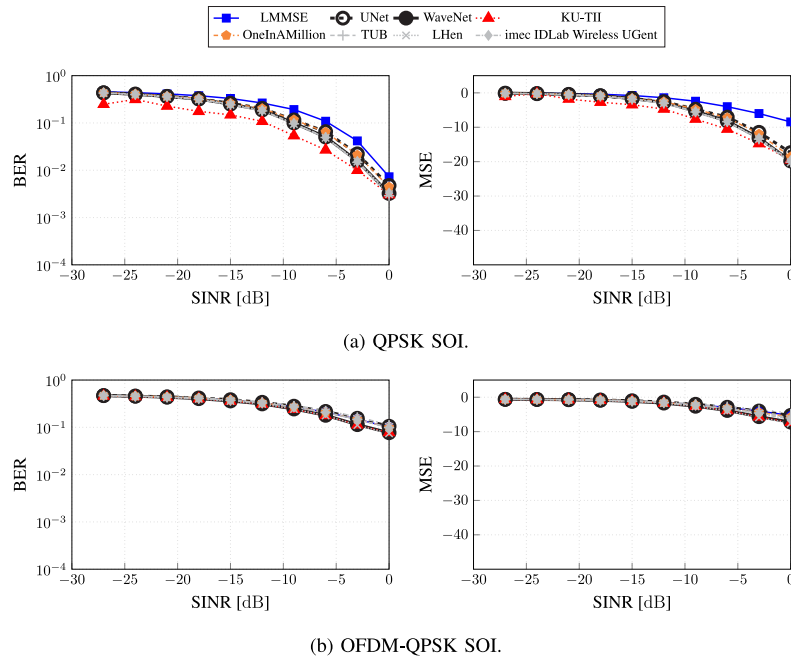
As we can see, there are some teams whose solutions improved upon the baselines in some cases involving EMISignal1, CommSignal2, and CommSignal5G1 (see Figs. 12(b), 13, and 15(a)). For example, "KU-TII" [35] especially shines in those mixture involving CommSignal2 interference, where they gain more than an order of magnitude in BER at SINR values below −20 dB compared to our baseline architectures, and "OneInAMillion" [36] performs especially well in the QPSK + CommSignal5G mixture, where they almost achieved an order of magnitude gain in BER at SINR values below −20 dB. Conversely, mixtures with CommSignal3 consistently challenge all methods. While we believe it is a multicarrier signal with a high data rate, the specific reasons for the difficulty to separate CommSignal3 from the SOI remain unclear, warranting further investigation. We note that, at least from a theoretical perspective, it should come as no surprise that one architecture is superior to others with respect to one fidelity measure (e.g., MSE), and is no longer superior with respect to a different one (e.g., minimum error probability). Indeed, a DNN is trained for a specific goal, via choosing

---

[14]The bit regression baseline code is available at https://github.com/RFChallenge/rfchallenge_singlechannel_starter/tree/main/example/demod_bitregression.

(a) QPSK SOI.

(b) OFDM-QPSK SOI.

FIGURE 13. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with CommSignal2 interference. This figure includes the LMMSE benchmark, along with our proposed data-driven methods and the proposed methods of the participants of the ICASSP'24 SP Grand Challenge.
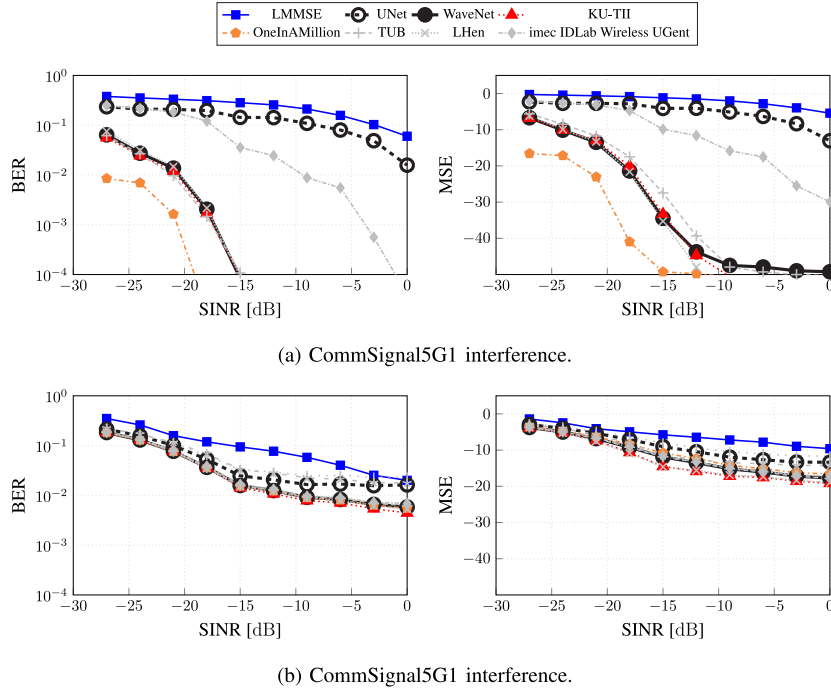

(a) QPSK SOI.

(b) OFDM-QPSK SOI.

FIGURE 14. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with CommSignal3 interference. This figure includes the LMMSE benchmark, along with our proposed data-driven methods and the proposed methods of the participants of the ICASSP'24 SP Grand Challenge.

a single objective function, and therefore cannot necessarily be optimal in more than one sense.[15]

These results show the potential of data-driven, deep-learning-based solutions to provide significant improvements in interference rejection tasks when the interference has unknown structures that can be learned. They also demonstrate that innovative solutions are needed to achieve further performance gains for challenging signals such as CommSignal3.

---

[15]However, this may happen in special cases, such as the MMSE and maximum *a posteriori* (MAP) estimators in the Gaussian signal model.

(a) CommSignal5G1 interference.

(b) CommSignal5G1 interference.

FIGURE 15. BER and MSE as a function of the target SINR for QPSK and OFDM-QPSK SOI, with CommSignal5G1 interference. This figure includes the LMMSE benchmark, along with our proposed data-driven methods and the proposed methods of the participants of the ICASSP'24 SP Grand Challenge.

V. FUTURE DIRECTIONS AND CONCLUDING REMARKS

In this section, we explore potential future research directions that can further advance the field of data-driven source separation of RF signals using deep learning techniques. We end the section with concluding remarks, encapsulating the key findings of this paper.

A. FUTURE RESEARCH DIRECTIONS

As demonstrated in the previous section—for a variety of signals using several methods—data-driven deep learning algorithms for RF source separation can yield significant performance gains. However, to make these techniques practically relevant, many aspects of this problem remain to be explored. Below, we outline and briefly discuss some of the more important ones.

One natural extension of this work is to investigate scenarios where the generation process of the SOI is *unknown*. Such cases are more challenging, as there is less prior information to exploit when designing the overall separation algorithm. For example, the MSE results obtained in Section IV can potentially be further improved if we leverage knowledge of the modulation scheme of the SOI via the decoding processing chain subsequent to the DNN-based interference mitigation module. Clearly, when the signal generation process is unknown, this approach is not applicable, and more sophisticated DNN architectures would be required to achieve the same performance.

As a preliminary empirical study on this front, we have used CommSignal2 as the SOI, and evaluated the MSE using the UNet and WaveNet architectures (Section III), as well as LMMSE (Section III-A). We used the remaining signals in the dataset as interference signals. The results, presented in Fig. 16, demonstrate that both the UNet and the WaveNet architectures outperforms LMMSE, except at low SINR levels with CommSignal3 as the interference.

Yet another extension to a prevalent setting is to multiple-input multiple-output (MIMO) systems, which are a cornerstone of modern wireless communication standards such as 5G and Wi-Fi. In this case, nontrivial spatial patterns can potentially be learned and exploited. Although the multi-channel source separation literature is significantly richer than the single-channel one (e.g., [43], [44], [45], [46]), the discussion revolving *data-driven deep-learning-based* methods in this context has not yet been comparatively addressed. Exploring how to utilize such spatial patterns *in tandem* with the signals' unique statistical characteristics, including temporal structures, constitutes a promising avenue for future research.

Other important extensions refer to potential physical effects induced by the channel. For example, a more comprehensive signal model may incorporate effects such as arbitrary time shifts and fading. More generally, an extended model of (1) can be expressed as:

$$\mathsf{y}[n] = \mathcal{H}\{\mathsf{s}[n]\} + \mathsf{b}\big[n - \mathsf{k}_b\big], \ n \in \mathbb{Z}, \qquad (6)$$

where, as before, $\mathsf{s}[n]$ and $\mathsf{b}[n]$ are the SOI and interference, respectively, and $\mathcal{H}\{\cdot\}$ denotes the channel response. For example, the channel could be $\mathcal{H}\{\mathsf{s}[n]\} = \alpha[n] \cdot \mathsf{s}[n - \mathsf{k}_s]$ for some time-varying fading coefficient coefficient $\alpha[n] \in \mathbb{C}$ and an unknow delay $\mathsf{k}_s \in \mathbb{N}$ (e.g., [47]). In particular, $\mathcal{H}\{\cdot\}$

(a) EMISignal1 interference.

(b) CommSignal3 interference.

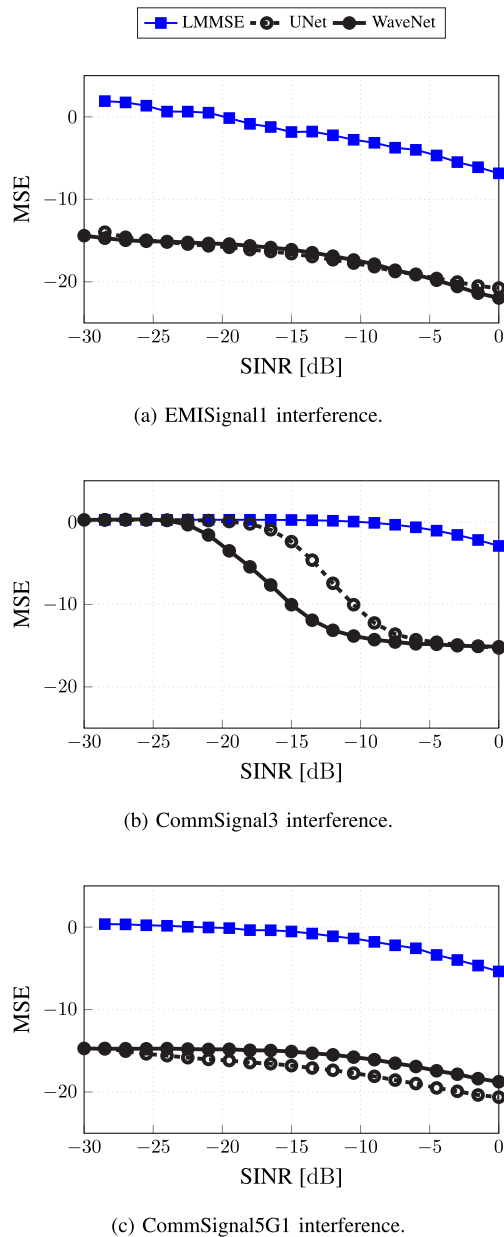(c) CommSignal5G1 interference.

FIGURE 16. MSE as a function of the target SINR for all combinations of CommSignal2 SOI and the rest of interference types considered in this work.

is not necessarily linear or time-invariant. The advantage of the data-driven approach is that it can potentially learn to compensate for nonlinear, time-varying effects, provided they are governed by some learnable statistical law, and are well captured in the available datasets for training.

A central attribute of the data-driven solution approach presented in this work is that each interference mitigation module, once trained, is tailored for specific mixtures of SOI and interference types. While this approach can lead to statistically optimal separation performance, its robustness is not guaranteed. Furthermore, it is highly demanding in terms of the system resources, such as memory, to maintain a separate NN for each mixture type, which can be

impractical in many scenarios. A desirable, more efficient, robust alternative would be to maintain a *single* model capable of mitigating *multiple* interference types. Such a model can establish the groundwork for developing a foundational RF signal separation model [48], [49], where signal characteristics (e.g., codes, modulations, pulse shaping) act as the "modalities".

Recent work [50] introduced a foundation model for wireless channels, using an encoder-decoder transformer trained to predict masked channel embeddings, successfully addressing downstream tasks like beam prediction and LOS/NLOS classification. Similarly, exploring a foundation model trained for interference rejection could unlock downstream tasks such as demodulation or constellation classification. A crucial first step in this direction is demonstrating that a single DNN can handle multiple interference types with performance comparable to specialized models, in terms of both inference time and separation fidelity.

B. CONCLUDING REMARKS

In this paper we illustrate the potential of deep learning-based methods for source separation of RF signals. Specifically, we show that mitigating strong unintentional interference from other RF emitting sources operating at the same time and the same frequency band with data-driven methods leads to considerable gains relative to traditional, "hand-crafted" methods. Through extensive simulation experiments, we demonstrate the superior performance of deep learning architectures, such as UNet and WaveNet, over the traditional signal processing methods of matched filtering and LMMSE estimation across various scenarios. Results from different leading research teams that participated in the "Data-Driven Radio Frequency Signal Separation Challenge", hosted as an ICASSP'24 Signal Processing Grand Challenge [27], show that further improving significantly beyond the established deep-learning benchmarks is nontrivial, especially in mixtures involving multi-carrier signals.

Ultimately, these results represent merely an initial phase of a more extensive journey towards integrating AI capabilities into receivers for enhanced interference rejection. The path forward would involve addressing additional, theoretical and practical, related problems, including interpretability of DNNs, and presenting viable solutions to demonstrate the tangible benefits of these approaches. Indeed, the results motivate further research and development for this dynamic domain within the broader community, with the ultimate goal of significantly improving future generations of RF systems spanning diverse applications.

REFERENCES

[1] "VR and AR pushing connectivity limits." Accessed: Feb. 26, 2024. [Online]. Available: https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/presentation_-_vr_and_ar_are_pushing_connectivity_limits_-web_0.pdf

[2] M. Hirzallah, W. Afifi, and M. Krunz, "Full-duplex-based rate/mode adaptation strategies for Wi-Fi/LTE-U coexistence: A POMDP approach," IEEE J. Sel. Areas Commun., vol. 35, no. 1, pp. 20–29, Nov. 2017.

[3] G. Naik, J.-M. Park, J. Ashdown, and W. Lehr, "Next generation Wi-Fi and 5G NR-U in the 6 GHz bands: Opportunities and challenges," IEEE Access, vol. 8, pp. 153027–153056, 2020.

[4] T. Oyedare, V. K. Shah, D. J. Jakubisin, and J. H. Reed, "Interference suppression using deep learning: Current approaches and open challenges," IEEE Access, vol. 10, pp. 66238–66266, 2022.

[5] H. L. V. Trees, Detection, Estimation, and Modulation Theory, Part I. New York, NY, USA: Wiley, 2001.

[6] T. Shilong, C. Shaohe, Z. Hui, and W. Jian, "Particle filtering based single-channel blind separation of co-frequency MPSK signals," in Proc. IEEE Int. Symp. Intell. Signal Process. Commun. Syst., Feb. 2007, pp. 582–585.

[7] T. Shilong, Z. Hui, and G. Na, "Single-channel blind separation of two QPSK signals using per-survivor processing," in IEEE Asia Pac. Conf. Circuits Syst. (APCCAS), Dec. 2008, pp. 473–476.

[8] J. Lee, D. Toumpakaris, and W. Yu, "Interference mitigation via joint detection," IEEE J. Sel. Areas Commun., vol. 29, no. 6, pp. 1172–1184, Jun. 2011.

[9] P. Chevalier, J.-P. Delmas, and M. Sadok, "Third-order volterra MVDR beamforming for non-Gaussian and potentially non-circular interference cancellation," IEEE Trans. Signal Process., vol. 66, no. 18, pp. 4766–4781, Jul. 2018.

[10] T. J. O'Shea and N. West, "Radio machine learning dataset generation with GNU radio," in Proc. GNU Radio Conf., vol. 1, no. 1, pp. 1–6, Sep. 2016.

[11] "RF challenge—AI accelerator," MIT RLE. Accessed: Dec. 3, 2024. [Online]. Available: https://rfchallenge.mit.edu

[12] D. Stoller, S. Ewert, and S. Dixon, "Wave-U-net: A multi-scale neural network for end-to-end audio source separation," in Proc. Int. Soc. Music Inf. Retrieval Conf., May 2018, pp. 334–340.

[13] A. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 24, no. 9, pp. 1652–1664, Sep. 2016.

[14] Y. Gandelsman, A. Shocher, and M. Irani, "'Double-DIP': Unsupervised image decomposition via coupled deep-image-priors," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 11026–11035.

[15] P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Joint optimization of masks and deep recurrent neural networks for monaural source separation," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 23, no. 12, pp. 2136–2147, Dec. 2015.

[16] T. Jayashankar, G. C. F. Lee, A. Lancho, A. Weiss, Y. Polyanskiy, and G. W. Wornell, "Score-based source separation with applications to digital communication signals," in Proc. Adv. Neural Inform. Proc. Syst. (NeurIPS), New Orleans, LA, USA, Dec. 2023, pp. 1–34. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/106b2434b8d496c6aed9235d478678af-Paper-Conference.pdf

[17] G. C. F. Lee, A. Weiss, A. Lancho, Y. Polyanskiy, and G. W. Wornell, "On neural architectures for deep learning-based source separation of co-channel OFDM signals," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Jun. 2023, pp. 1–5.

[18] "Single-channel source separation: Preliminary test of neural network architectures." MIT RLE: RF Challenge. 2022. [Online]. Available: https://github.com/RFChallenge/SCSS_DNN_Comparison

[19] T. R. Oyedare, "A comprehensive analysis of deep learning for interference suppression, sample and model complexity in wireless systems," Ph.D. dissertation, Dept. Electr. Eng., Virginia Polytechnic Inst., State Univ., Blacksburg, VA, USA, Mar. 2024.

[20] "RF datasets for machine learning," DeepSig. Accessed: Dec. 3, 2024. [Online]. Available: https://www.deepsig.ai/datasets

[21] "A Web-based SDR toolkit for analyzing, processing, and sharing RF recordings." IQ Engine. Accessed: Dec. 3, 2024. [Online]. Available: https://iqengine.org/

[22] G. C. F. Lee et al., "Exploiting temporal structures of cyclostationary signals for data-driven single-channel source separation," in Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP), Aug. 2022, pp. 1–6.

[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Aug. 2009, pp. 248–255.

[25] K. Cook, G. Grinstein, and M. Whiting, "The VAST challenge: History, scope, and outcomes: An introduction to the special issue," Inf. Visualization, vol. 13, no. 4 pp. 301–312, 2014.

[26] P. Luszczek et al., "Introduction to the HPC challenge benchmark suite," Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Rep. 860347, 2005.

[27] T. Jayashankar et al., "The data-driven radio frequency signal separation challenge," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Apr. 2024, pp. 53–54.

[28] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: A survey," IEEE Trans. Veh. Technol., vol. 58, no. 4, pp. 1673–1694, May 2009.

[29] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (Lecture Notes in Computer Science 9351), N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham, Switzerland: Springer, Nov. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[30] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, "Radio frequency interference mitigation using deep convolutional neural networks," Astron. Comput., vol. 18, pp. 35–39, Jan. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213133716301056

[31] E. Tzinis, Z. Wang, and P. Smaragdis, "Sudo RM-RF: Efficient networks for universal audio source separation," in Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP), Sep. 2020, pp. 1–6.

[32] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," IEEE Trans. Signal Process., vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[33] A. van den Oord et al., "WaveNet: A generative model for raw audio," Sep. 2016, arXiv:1609.03499.

[34] D. Rethage, J. Pons, and X. Serra, "A wavenet for speech denoising," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Apr. 2018, pp. 5069–5073.

[35] Y. Tian, A. Alhammadi, A. Quran, and A. S. Ali, "A novel approach to wavenet architecture for RF signal separation with learnable dilation and data augmentation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW), Apr. 2024, pp. 79–80.

[36] F. Damara, Z. Utkovski, and S. Stanczak, "Signal separation in radio spectrum using self-attention mechanism," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW), Apr. 2024, pp. 99–100.

[37] Z. Kong, W. Ping, A. Dantrey, and B. Catanzaro, "Speech denoising in the waveform domain with self-attention," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), May 2022, pp. 7867–7871.

[38] L. Henneke, "Improving data-driven RF signal separation with SOI-matched autoencoders," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW), Apr. 2024, pp. 45–46.

[39] Ç. Yapar et al., "DEMUCS for data-driven RF signal denoising," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW), Apr. 2024, pp. 95–96.

[40] A. Défossez, G. Synnaeve, and Y. Adi, "Real time speech enhancement in the waveform domain," in Proc. Interspeech, 2020, pp. 3291–3295.

[41] A. Défossez, N. Usunier, L. Bottou, and F. Bach, "Music source separation in the waveform domain," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021, pp. 1–16. [Online]. Available: https://openreview.net/forum?id=HJx7uJStPH

[42] M. Naseri, J. Fontaine, I. Moerman, E. De Poorter, and A. Shahid, "A U-net architecture for time-frequency interference signal separation of RF waveforms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW)*, Apr. 2024, pp. 91–92.

[43] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.

[44] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[45] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.

[46] A. Weiss and A. Yeredor, "A maximum likelihood-based minimum mean square error separation and estimation of stationary Gaussian sources from noisy mixtures," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 5032–5045, Oct. 2019.

[47] A. Lancho et al., "Data-driven blind synchronization and interference rejection for digital communication signals," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 2296–2302.

[48] R. Bommasani et al., "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.

[49] J. Fontaine, A. Shahid, and E. D. Poorter, "Towards a wireless physical-layer foundation model: Challenges and strategies," Feb. 2024, *arXiv:2403.12065*.

[50] S. Alikhani, G. Charan, and A. Alkhateeb, "Large wireless model (LWM): A foundation model for wireless channels," Nov. 2024, *arXiv:2411.08872*.

ALEJANDRO LANCHO (Member, IEEE) received the B.E., M.Sc., and Ph.D. degrees in electrical engineering from the Universidad Carlos III de Madrid, Spain, in 2013, 2014, and 2019, respectively. From 2019 to 2021, he was a Postdoctoral Researcher with the Chalmers University of Technology, Sweden. From 2021 to 2023, he was a Marie Curie Postdoctoral Global Fellow with the Massachusetts Institute of Technology, USA. Since October 2023, he has been with the Universidad Carlos III de Madrid, where he is currently an Atracción de Talento César Nombela Research Fellow. His research interests include information theory and deep learning for wireless communications. He was among the six finalists for the IEEE Jack Keil Wolf ISIT Student Paper Award at the 2017 IEEE International Symposium on Information Theory.
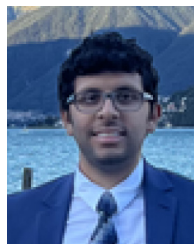
AMIR WEISS (Senior Member, IEEE) received the B.Sc. (magna cum laude), M.Sc., and Ph.D. degrees in electrical engineering from Tel Aviv University (TAU), Tel-Aviv, Israel, in 2013, 2015, and 2020, respectively. From 2019 to 2020, he was a Postdoctoral Fellow with the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. From 2020 to 2023, he was a Postdoctoral Associate with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. He has held a researcher position with Elbit Systems, EW and SIGINT Elisra Ltd., Holon, Israel, from 2013 to 2020, specializing in detection and estimation of RADAR and SONAR signals. Since 2024, he has been the Faculty with Bar-Ilan University, Ramat Gan, Israel, where he is a Senior Lecturer (Associate Professor equivalent) with the Faculty of Engineering. He is also the Head of AI and Technological Innovation, Medibyt Ltd., Israel. His main research areas are in statistical and digital signal processing, estimation theory, and machine learning, with applications to localization, communication, and compression. He received the scholarship for excellent M.Sc. students from the Faculty of Engineering, TAU in 2015, and The David and Paulina Trotsky foundation award for outstanding Ph.D. students in 2019. He is a recipient of the 2021 ICASSP Outstanding Paper Award. He also received the Nadav Levanon studies prize for graduate students in 2016, the scientific publication prize (three times) in 2017, 2018, and 2020, respectively, and the David Burshtein scientific publication prize in 2019, all from the The Yitzhak and Chaya Weinstein Research Institute for Signal Processing. He is a recipient of the 2024 Alon Scholarship for the Integration of Outstanding Faculty.

GARY C. F. LEE (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2019 and 2023, respectively. He is a Research Scientist with the Institute for Infocomm Research (A*STAR). His research focuses on artificial intelligence and machine learning (AI/ML) for next-generation wireless communications systems, with interests spanning the areas of signal processing, communications engineering, and AI/ML.

TEJAS JAYASHANKAR (Student Member, IEEE) received the B.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 2019 and the M.S. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT) in 2022, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science where is advised by Prof. Wornell and is supported by the MIT Claude E. Shannon Fellowship. His research interests lie at the intersection of statistical inference, generative modeling and signal processing. His more recent research focus is centered around score estimation, diffusion models, and efficient neural synthesis.

BINOY G. KURIEN received the master's and Ph.D. degrees in engineering sciences from the School of Engineering and Applied Sciences, Harvard University in 2009 and 2016, respectively. He is currently an Assistant Group Leader with the Tactical Edge Communications Group, MIT Lincoln Laboratory, Lexington, MA, USA. His research interests are in dynamic spectrum access, applications of machine learning to wireless communications, array signal processing, and remote sensing.

YURY POLYANSKIY (Fellow, IEEE) received the M.S. degree in applied mathematics and physics from the Moscow Institute of Physics and Technology, Moscow, Russia, in 2005 and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2010. He is a Cutten Professor of electrical engineering and computer science, a member of IDSS and LIDS with MIT. His research interests span information theory, machine learning and statistics. He won the 2020 IEEE Information Theory Society James Massey Award, the 2013 NSF CAREER award, and the 2011 IEEE Information Theory Society Paper Award.

GREGORY W. WORNELL (Fellow, IEEE) received the B.A.Sc. degree in electrical engineering and computer science from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT) in 1985, 1987, and 1991, respectively.

Since 1991 he has been a Faculty with MIT, where he is the Sumitomo Professor of Engineering with the Department of Electrical Engineering and Computer Science (EECS) and the Schwarzman College of Computing, and an Area Co-Chair of the EECS Doctoral Program. At MIT he leads the Signals, Information, and Algorithms Laboratory, and is affiliated with the Research Laboratory of Electronics, the Computer Science and Artificial Intelligence Laboratory, and the Institute for Data, Systems and Society. He has held visiting appointments with the Department of Electrical Engineering and Computer Science at the University of California, Berkeley, CA, USA, from 1999 to 2000, the Hewlett-Packard Laboratories, Palo Alto, CA, USA, in 1999, and the AT&T Bell Laboratories, Murray Hill, NJ, USA, from 1992 to 1993. His research interests and publications span the areas of signal processing, information theory, statistical inference, artificial intelligence, and information security, and include architectures for sensing, learning, computing, communication, and storage; systems for computational imaging, vision, and perception; aspects of computational biology and neuroscience; and the design of wireless networks. He has been involved in the Information Theory and Signal Processing societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities.

Prof. Wornell has won a number of awards for both his research and teaching, including the 2019 IEEE Leon K. Kirchmayer Graduate Teaching Award.