Data-Driven Blind Synchronization and Interference Rejection for Digital Communication Signals

Alejandro Lancho*, Amir Weiss*, Gary C.F. Lee,
Jennifer Tang, Yuheng Bu, Yury Polyanskiy, and Gregory W. Wornell
Massachusetts Institute of Technology, Cambridge, MA, USA
Emails: {lancho, amirwei, glcf411, jstang, buyuheng, ypol, gww}@mit.edu

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001513

***Abstract*—We study the potential of data-driven deep learning methods for separation of two communication signals from an observation of their mixture. In particular, we assume knowledge on the generation process of one of the signals, dubbed signal of interest (SOI), and no knowledge on the generation process of the second signal, referred to as interference. This form of the single-channel source separation problem is also referred to as interference rejection. We show that capturing high-resolution temporal structures (nonstationarities), which enables accurate synchronization to both the SOI and the interference, leads to substantial performance gains. With this key insight, we propose a domain-informed neural network (NN) design that is able to improve upon both "off-the-shelf" NNs and classical detection and interference rejection methods, as demonstrated in our simulations. Our findings highlight the key role communication-specific domain knowledge plays in the development of data-driven approaches that hold the promise of unprecedented gains.**

***Index Terms*—Blind synchronization, source separation, interference rejection, deep neural network, supervised learning.**

I. Introduction

The proliferation of wireless devices is leading to an increasingly crowded radio spectrum, and consequently, spectrum sharing will be unavoidable [1], [2]. Thus, different wireless communication systems will coexist in the same frequency bands, thereby generating unintentional interferences among them. In order to maintain high reliability, separation of the overlapping signals from the received mixture will become an essential building block in such communication systems.

In the image and audio domains, machine learning techniques have been successfully applied for *source separation*, e.g., [3]. These methods usually exploit domain knowledge relating to the signals' structures. For example, color features and local dependencies are useful for separating natural images [4], whereas time-frequency spectrogram masking methods are typically adopted for separating audio signals [5].

For communication signals, if the sources are separable in time and/or frequency, one can separate them via appropriate masking and classical filtering methods (see, e.g., [6]). The key challenge in this domain is the separation of signals overlapping in *both* time and frequency when the receiver is equipped with a single antenna, which inherently implies there is no spatial diversity to be exploited. This problem is also referred to as *single-channel* source separation (SCSS). In this case, standard approaches exploiting spatial diversity for blind source separation, such as [7], [8], are irrelevant.

Various methods are available in the literature to perform SCSS of digital communication signals. A common approach is maximum likelihood sequence estimation of the target signal, for which algorithms such as particle filtering [9] and per-surviving processing algorithms [10] can be used. However, such methods require prior knowledge of the signal models, which in practice may not be known or available.

Perhaps a more realistic approach 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., [11]), allowing for a *data-driven* approach. In this setup, deep neural networks (DNNs) arise as a natural choice. This problem has been recently promoted by the "RF Challenge" [12].

In this paper, we study the data-driven SCSS problem where two communication signals overlap in time and frequency, and the receiver is equipped with one single antenna. We consider a signal of interest (SOI) whose generation process is known, and an interference signal with cyclic statistical properties that are unknown *a priori*—as is the case in standard protocols.[1] This problem is also referred to as *interference rejection*. As a performance measure, we consider the bit error rate (BER).

Contributions: We show that temporal nonstationarities of the signals constitute strong regularities that translate to better separation conditions. In particular, when such temporal structures exist, the notion of (time-)synchronization becomes not only sensible, but advantageous for separation. Based on our theoretical results that bind synchronization with MMSE optimal separation, we propose a data-driven DNN approach that is BER-superior to the classical methods of demodula-

Research was sponsored by the United States Air Force Research Laboratory and the United States Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. Alejandro Lancho has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101024432. G. C.F. Lee is supported by the National Science Scholarship from the Agency for Science, Technology and Research (A*STAR). This work is also supported by the National Science Foundation under Grant No CCF-2131115. *These authors contributed equally to this work.

[1] We only assume the cyclic period is known. In practice, provided a dataset of the respective signal, this parameter can be consistently estimated [13].

tion with matched filtering (MF) and interference rejection with linear minimum mean-square error (LMMSE) estimation of the SOI. Our proposed DNNs architectures, which can incorporate explicit synchronization, are inspired by specific domain knowledge, relevant to digital communication signals.

Notation: We use lowercase letters with standard font and sans-serif font, e.g., x and x, to denote deterministic and random scalars, respectively. Similarly, we use $\mathbf{x}$ and $\mathbf{x}$ for deterministic and random vectors, respectively; and $\mathbf{X}$ and $\mathbf{X}$ for deterministic and random matrices, respectively. The uniform distribution over a set $\mathcal{S}$ is denoted as $\mathrm{Unif}(\mathcal{S})$, and for $K \in \mathbb{N}$, we denote $\mathcal{S}_K \triangleq \{1, \ldots, K\}$. For brevity, we refer to the complex normal distribution as Gaussian. We denote $\mathbf{C}_{zw} \triangleq \mathbb{E}\left[\mathbf{z}\mathbf{w}^{\mathsf{H}}\right] \in \mathbb{C}^{N_z \times N_w}$ as the covariance matrix of $\mathbf{z} \in \mathbb{C}^{N_z \times 1}$ and $\mathbf{w} \in \mathbb{C}^{N_w \times 1}$ (specializing to $\mathbf{C}_{zz}$ for $\mathbf{z} = \mathbf{w}$).

II. Problem Formulation

We consider the single-channel, baseband signal model of a noisy mixture of two sources, given by

$$\mathsf{y}[n] = \mathsf{s}[n - \mathsf{k}_s] + \rho_{\text{SIR}}^{-1/2}\mathsf{b}[n - \mathsf{k}_b] + \rho_{\text{SNR}}^{-1/2}\mathsf{w}[n], \ n \in \mathbb{Z}, \quad (1)$$

where $\mathsf{s}[n], \mathsf{b}[n] \in \mathbb{C}$ are assumed to be cyclostationary processes with known fundamental cyclic periods $K_s, K_b \in \mathbb{N}$, respectively; $\mathsf{w}[n] \in \mathbb{C}$ denotes additive white Gaussian noise, statistically independent of $\mathsf{s}[n]$ and $\mathsf{b}[n]$; and $\rho_{\text{SIR}}, \rho_{\text{SNR}} \in \mathbb{R}_+$. We refer to the signal $\mathsf{s}[n]$ as the SOI, and to $\mathsf{b}[n]$ as interference. The variables $\mathsf{k}_s, \mathsf{k}_b \in \mathbb{Z}$ denote unknown (discrete) time-shifts with respect to the start of the cyclic periods of $\mathsf{s}[n]$ and $\mathsf{b}[n]$, respectively, where the start of the cyclic periods are chosen arbitrarily to be at $n = 0$ without loss of generality. Hence, we assume that $\mathsf{k}_s \sim \mathrm{Unif}\left(\mathcal{S}_{K_s}\right)$ and $\mathsf{k}_b \sim \mathrm{Unif}\left(\mathcal{S}_{K_b}\right)$.

Let $\mathbf{y} \triangleq [\mathsf{y}[1] \cdots \mathsf{y}[N]]^{\mathrm{T}}$, $\mathbf{s}(\mathsf{k}_s) \triangleq [\mathsf{s}[1 - \mathsf{k}_s] \cdots \mathsf{s}[N - \mathsf{k}_s]]^{\mathrm{T}}$, $\mathbf{b}(\mathsf{k}_b) \triangleq [\mathsf{b}[1 - \mathsf{k}_b] \cdots \mathsf{b}[N - \mathsf{k}_b]]^{\mathrm{T}}$, and $\mathbf{w} \triangleq [\mathsf{w}[1] \cdots \mathsf{w}[N]]^{\mathrm{T}}$. Then, we may compactly write (1) for N samples as

$$\mathbf{y} = \mathbf{s}(\mathsf{k}_s) + \rho_{\text{SIR}}^{-1/2}\mathbf{b}(\mathsf{k}_b) + \rho_{\text{SNR}}^{-1/2}\mathbf{w} \in \mathbb{C}^{N \times 1}. \quad (2)$$

We further assume that $\mathbf{s}(\mathsf{k}_s)$ and $\mathbf{b}(\mathsf{k}_b)$ are statistically independent, which is a reasonable assumption in scenarios of unintentional interference, for which each source is not actively jamming or adapting to the other signals present in the environment. For simplicity of the exposition, we assume that $\mathbf{s}(\mathsf{k}_s)$ and $\mathbf{b}(\mathsf{k}_b)$ are zero-mean, unit-average-power, i.e., their (possibly time-varying) variance averages to 1. In this case, the parameters $\rho_{\text{SIR}}, \rho_{\text{SNR}}$ represent the signal-to-interference ratio (SIR) and signal-to-noise ratio (SNR) at the receiver, respectively.

The goal is to produce an estimate of $\mathbf{s}(\mathsf{k}_s)$ from $\mathbf{y}$, denoted by $\widehat{\mathbf{s}}$, so that given some metric ℓ, the cost $\mathbb{E}[\ell(\widehat{\mathbf{s}}, \mathbf{s}(\mathsf{k}_s))]$ is minimized. This problem is referred to as SCSS.

As mentioned in Section I, we assume we do not have precise knowledge of the underlying distributions of the SOI and the interference. However, we assume the availability of a dataset of the signals and their respective time-shifts $(\mathbf{s}(\mathsf{k}_s), \mathsf{k}_s)$ and $(\mathbf{b}(\mathsf{k}_b), \mathsf{k}_b)$, allowing for a data-driven approach. Examples of such datasets can be found in [12], [14].

III. The Gain in Synchronization to Interference

Before we present our approach to the SCSS problem formulated in Section II, we provide an analysis of an asymptotically optimal estimator of $\mathbf{s}(\mathsf{k}_s)$ for the metric $\ell(\mathbf{x}, \mathbf{z}) \triangleq \|\mathbf{x} - \mathbf{z}\|_2^2$, which will shed light on key aspects in optimal separation and the role of *synchronization to interference*.

In this section, we assume that $\mathsf{s}[n]$ and $\mathsf{b}[n]$ are Gaussian processes, which is a reasonable assumption to model some communication signals, e.g., [15]. In this case, we define[2]

$$\mathsf{v}[n - \mathsf{k}_b] \triangleq \rho_{\text{SIR}}^{-1/2}\mathsf{b}[n - \mathsf{k}_b] + \rho_{\text{SNR}}^{-1/2}\mathsf{w}[n], \quad n \in \mathbb{Z}, \quad (3)$$

such that $\mathbf{v}(\mathsf{k}_b) \triangleq [\mathsf{v}[1 - \mathsf{k}_b] \cdots \mathsf{v}[N - \mathsf{k}_b]]^{\mathsf{T}} \in \mathbb{C}^{N \times 1}$ is the "equivalent noise", which, given k_b, is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{C}_{vv})$. Thus, (2) simplifies to

$$\mathbf{y} = \mathbf{s}(\mathsf{k}_s) + \mathbf{v}(\mathsf{k}_b) \in \mathbb{C}^{N \times 1}. \quad (4)$$

Note that, generally, the equivalent noise term $\mathbf{v}(\mathsf{k}_b)$ is not temporally white (as opposed to $\mathbf{w}$), and exhibits a potentially informative statistical structure (e.g., in the form of $\mathbf{C}_{vv}$) that can be exploited for enhanced separation performance.

A. Linear minimum mean-square error (MMSE) Estimation

A computationally attractive approach, which already exploits (some of) the underlying statistics of both of the components of the mixture (4), is optimal *linear* estimation. The LMMSE estimator [16], given by (assuming $\det(\mathbf{C}_{yy}) \neq 0$)

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

is constructed using the statistics of the mixture that inherently takes into account the potentially non-trivial structure of $\mathbf{C}_{vv}$, i.e., some form of deviation from a scaled identity matrix.

However, while (5) coincides with the MMSE estimator for jointly Gaussian processes, it is generally suboptimal due to the linearity constraint. Specifically, in our case, although the processes $\mathsf{s}[n], \mathsf{v}[n]$ are jointly Gaussian, $\mathbf{s}(\mathsf{k}_s)$ and $\mathbf{v}(\mathsf{k}_b)$ are *not* even marginally Gaussian. Indeed, $\mathbf{s}(\mathsf{k}_s)$ and $\mathbf{v}(\mathsf{k}_b)$ are Gaussian mixtures due to the random time-shifts $\mathsf{k}_s, \mathsf{k}_b$. It then follows that (5) is in fact not optimal, as shown next.

B. MMSE Estimation

The optimal estimator in the MMSE sense is known to be the conditional expectation,

$$\widehat{\mathbf{s}}_{\text{MMSE}} \triangleq \mathbb{E}[\mathbf{s}(\mathsf{k}_s)|\mathbf{y}] \in \mathbb{C}^{N \times 1}, \quad (6)$$

whose mean-squared error (MSE) is an achievable lower bound of the MSE of *any* estimator of $\mathbf{s}(\mathsf{k}_s)$. However, in most practical cases, (6) is hard to obtain analytically and computationally. In our case, by using the law of total expectation in (6), the MMSE estimator is given by the more explicit and convenient form

$$\begin{aligned}\widehat{\mathbf{s}}_{\text{MMSE}} &= \mathbb{E}\left[\mathbb{E}[\mathbf{s}(\mathsf{k}_s)|\mathbf{y}, \mathsf{k}_s, \mathsf{k}_b]|\mathbf{y}\right] \overset{(\star)}{=} \mathbb{E}\left[\widehat{\mathbf{s}}_{\text{LMMSE}}(\mathsf{k}_s, \mathsf{k}_b)|\mathbf{y}\right] \\ &= \sum_{m_s=1}^{K_s}\sum_{m_b=1}^{K_b}\mathbb{P}[\mathsf{k}_s = m_s, \mathsf{k}_b = m_b|\mathbf{y}]\,\widehat{\mathbf{s}}_{\text{LMMSE}}(m_s, m_b), \quad (7)\end{aligned}$$

[2]Since $\mathsf{w}[n]$ is white (and therefore stationary), $\mathsf{w}[n - \mathsf{k}_b]$ is also white, hence without loss of generality we may indeed define (3) with the shift k_b.

where in $(\star)$ we have used the fact that, given the time-shifts, $\mathbf{s}(\mathsf{k}_s)$ and $\mathbf{y}$ are jointly Gaussian, and where $\widehat{\mathbf{s}}_{\text{LMMSE}}(m_s, m_b) \triangleq \mathbf{C}_{ss}(m_s)\left[\mathbf{C}_{ss}(m_s) + \mathbf{C}_{vv}(m_b)\right]^{-1}\mathbf{y}$, with

$$\mathbf{C}_{ss}(m) \triangleq \mathbb{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}|\mathsf{k}_s = m],\ \mathbf{C}_{vv}(m) \triangleq \mathbb{E}[\mathbf{v}\mathbf{v}^{\mathsf{H}}|\mathsf{k}_b = m]. \quad (8)$$

Put simply, (7) is a weighted average of $K_s \times K_b$ linear estimators, with the posterior probabilities—which are *non-linear* functions of the data $\mathbf{y}$—serving as the normalized weights. Even before taking into account the computation of the posteriors, the sum in (7) scales with the product of possible time-shifts $K_s \times K_b$, rendering $\widehat{\mathbf{s}}_{\text{MMSE}}$ often impractical.

As can be seen from (7), synchronization (i.e., knowledge of the time-shifts) already substantially simplifies the computation, since, in that case, only the (conditional) linear estimator $\widehat{\mathbf{s}}_{\text{LMMSE}}(m_s, m_b)$ is required. In other words, eliminating this type of randomness from the mixture $\mathbf{y}$ grants us lower computational complexity and a simple form of a linear estimator. Fortunately, a two-step "synchronization-separation" estimator can approach the MMSE estimator, thus enjoying asymptotic optimality at a substantially reduced computational burden.

To show this rigorously, for simplicity of the exposition, we assume hereafter (unless stated otherwise) that the receiver is synchronized to the SOI,[3] namely, $\mathsf{k}_s = 0$ and known. However, the result below can be generalized to the case where the SOI's time-shift k_s is random and unknown. Let

$$\widehat{\mathsf{k}}_b^{\text{MAP}} \triangleq \underset{m \in \mathcal{S}_{K_b}}{\arg\max}\, \mathbb{P}[\mathsf{k}_b = m|\mathbf{y}] \quad (9)$$

be the maximum *a posteriori* (MAP) estimator of k_s, and define the (suboptimal) "plug-in", MAP-based *quasi*-linear MMSE estimator

$$\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} \triangleq \widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{MAP}}) \in \mathbb{C}^{N\times 1}, \quad (10)$$

where, for brevity, we use $\widehat{\mathbf{s}}_{\text{LMMSE}}(m)$ to denote $\widehat{\mathbf{s}}_{\text{LMMSE}}(0, m)$. Furthermore, we define the MSEs, as a function of N, as

$$\varepsilon^2_{\text{MMSE}}(N) \triangleq \mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MMSE}} - \mathbf{s}\|_2^2\right] \in \mathbb{R}_+, \quad (11)$$

$$\varepsilon^2_{\text{MAP-QLMMSE}}(N) \triangleq \mathbb{E}[\|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} - \mathbf{s}\|_2^2] \in \mathbb{R}_+. \quad (12)$$

We now introduce a "temporal-diversity" condition (TDC) under which optimal synchronization is increasingly accurate.

Definition 1 (TDC): Let $\psi_N(\mathbf{y}, k) \triangleq \frac{1}{N}\mathbf{y}^{\mathsf{H}}\mathbf{C}_{yy}^{-1}(k)\mathbf{y} - 1$. The (sufficient) TDC is satisfied if there does not exist $k \in \mathcal{S}_{K_b}\backslash \mathsf{k}_b$ such that $\lim_{N\to\infty}|\psi_N(\mathbf{y}, k)| = 0$.

Lemma 1: Under the TDC, for any finite $\alpha \in \mathbb{R}_+$,

$$\mathbb{P}\left[\widehat{\mathsf{k}}_b^{\text{MAP}} \neq \mathsf{k}_b\right] = o\left(N^{-\alpha}\right). \quad (13)$$

Proof: See Appendix A. ∎

The theorem below shows that the two-step synchronization-separation approach (10) is asymptotically optimal.

Theorem 1: Under the TDC, we have

$$\lim_{N\to\infty} \frac{\varepsilon^2_{\text{MMSE}}(N)}{\varepsilon^2_{\text{MAP-QLMMSE}}(N)} = 1. \quad (14)$$

Proof: See Appendix B. ∎

[3]This is a reasonable assumption in most communication systems [17].

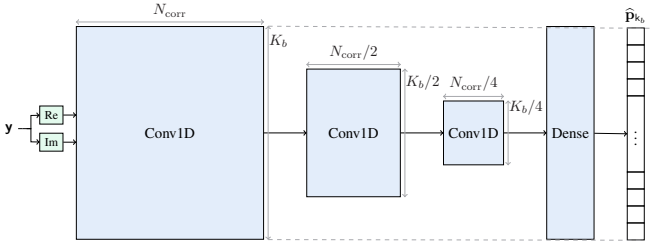

Fig. 1: Architecture of the proposed CNN for synchronization.

In words, Theorem 1 tells us that, when the time-shift can be uniquely detectable, first optimally synchronizing to the interference, and then using a suboptimal, quasi-linear estimator, is asymptotically equivalent to MMSE estimation. Further intuition to this type of behaviour, for maximum-likelihood-based MMSE estimation, is given in [8, Fig. 1].

C. Synchronization via convolutional neural networks (CNNs)

Although the estimator (10) is attractive in terms of its MSE performance, it nevertheless requires—both for synchronization and separation—*precise* knowledge of the underlying statistics, including the SIR and SNR, which can be hard to obtain in practice. In particular, without these statistics, it is impossible to obtain $\widehat{\mathsf{k}}_b^{\text{MAP}}$. However, when by measurement or generation, sufficiently large datasets with realizations of $\mathbf{s}(\mathsf{k}_s)$ and $\mathbf{b}(\mathsf{k}_b)$ are available, a data-driven approach can be taken.

To this end, we leverage the strong capabilities of CNNs for capturing intricate temporal structures, to train a synchronizer in a data-driven manner. Specifically, we propose the CNN-based architecture depicted in Fig. 1, which is trained in a supervised manner based on a labeled dataset of mixtures and the underlying interference time-shifts, $\{(\mathbf{y}^{(i)}, \mathsf{k}_b^{(i)}) : i \in \mathcal{S}_{I_T}\}$, where I_T is the size of the training dataset. We use a sufficiently large kernel size in the convolutional layers, which is proportional to the "effective correlation length"—denoted as N_{corr} in Fig. 1—so as to be able to capture the strongest, most informative temporal structures for estimation. Since the cyclic period K_b is assumed to be known, we train a model using the cross-entropy loss, which receives as its input the mixture $\mathbf{y}$ and outputs a vector of probabilities, denoted by $\widehat{\mathbf{p}}_{\mathsf{k}_b} \in [0,1]^{K_b\times 1}$. At inference time, we synchronize to the interference via $\widehat{\mathsf{k}}_b^{\text{CNN}} \triangleq \arg\max_{m\in\mathcal{S}_{K_b}} \mathbf{e}_m^{\mathsf{T}}\widehat{\mathbf{p}}_{\mathsf{k}_b}$ (cf. $\widehat{\mathsf{k}}_b^{\text{MAP}}$ in (9)), where $\mathbf{e}_m \in \mathbb{R}^{K_b\times 1}$ denotes the m-th standard basis vector.

In Fig. 2, we show the MSE for $\widehat{\mathbf{s}}_{\text{LMMSE}}, \widehat{\mathbf{s}}_{\text{MMSE}}$ and

$$\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}} \triangleq \widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{CNN}}), \quad (15)$$

for the same communication waveforms described in detail in Section V, but considering here Gaussian alphabets instead of the discrete and finite alphabets used in Section V. As seen, the linearity restriction (5) costs a considerable price in terms of the compromised performance relative to the lower bound, given by the MMSE. It is also evident that the MSE of the CNN-based quasilinear MMSE (QLMMSE) estimator $\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}}$ coincides with (11), which asymptotically coincides with the MAP-QLMMSE (12) by virtue of Theorem 1.

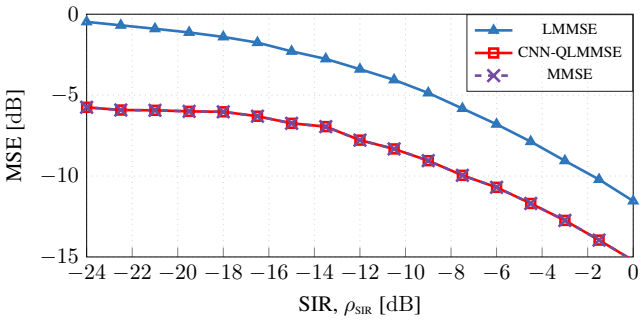

Fig. 2: MSE as a function of the SIR (ρ_{SIR}) for a fixed SNR (ρ_{SNR}) of 20 dB.

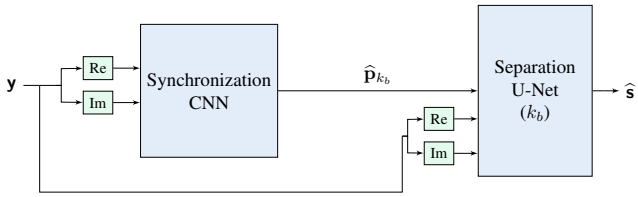

Fig. 3: System architecture of the DNN–based approach with an explicit CNN-based synchronization block prior to the separation block (U-Net).

All the above motivates our solution approach, and provides the theoretical foundations (as well as intuition) based on which we develop our system architecture, presented next.

IV. Interference Rejection via DNNs

We now present two supervised learning approaches for SCSS, used in this work as interference rejection methods. The first DNN architecture, depicted in Fig. 3, consists of two main building blocks: (i) CNN to perform synchronization to the interference, (ii) DNN (U-Net) to perform SCSS.[4] The key motivation to perform explicit synchronization is twofold. First, as explained in Section III-B, due to Theorem 1, explicit consistent synchronization decoupled from separation, although suboptimal, can asymptotically (as $N \to \infty$) lead to optimal separation with reduced complexity. Second, although a sufficiently rich DNN might be able to perform the synchronization and separation tasks jointly, for a given architecture, acquiring synchronization knowledge explicitly helps by reducing the complexity of the separation task. In Section V-A, we show that this decoupled approach can indeed lead to performance gains. However, Lemma 1 shows that there exists a realizable synchronization method that becomes increasingly accurate as the input size grows. While this can be exploited for explicit synchronization (e.g., Fig. 3), it could also imply that, under certain conditions, a DNN architecture would be able to "implicitly synchronize" and separate, namely superior performance would be achieved *without* explicit synchronization. This is shown in Section V-B.

The synchronization block is based on the CNN described in Section III-C (Fig. 1). The DNN for separation is based on the so-called U-Net (see Fig. 4) [18], which has some properties that makes it suitable to the specific informative features of digital communication signals. In particular, its CNN building blocks allow us to input and process long time intervals (e.g., $N > 10^4$), which cannot be processed using classical methods. In turn, processing such long signals allows for exploitation of temporal structures on a different scale, which can (and does) lead to substantial performance gains.

As shown in Fig. 4, our DNN approach departs from standard implementations intended to deal with images (2D signals). To handle 1D complex-valued, time-series communication signals, we use 1D convolutional layers. Furthermore, differently from standard CNN-based architectures that are designed to deal with images and hence use short kernels of size ~ 3 in all layers, our U-Net architecture utilizes a sufficiently long kernel in the first convolutional layer (denoted by κ in Fig. 4). This enables to capture the most influential temporal structures of the SOI and interference, which can lead to an order of magnitude gains, as demonstrated below.

For training, we input the stacked real and imaginary parts of $\mathbf{y}$ as separate channels to both the synchronization-to-interference CNN and the separation U-Net. For separation, if explicit synchronization is performed, we mimic a non-linear version of (10) by using an instance of the DNN architecture depicted in Fig. 4 for each possible output of the synchronization-to-interference CNN block. In other words, we implement a "conditional separation" block for each possible time-shift of the interference. If explicit synchronization-to-interference is not used, the raw unprocessed mixture is (always) fed into to the same DNN separation block.

The training set is processed as such to yield a labeled dataset (mixture $\mathbf{y}$ and ground-truth reference signal $\mathbf{s}$). As a loss function, we use the empirical MSE. For full implementation details, see our Github repository.[5]

V. Numerical Results

We generate synthetic mixtures $\mathbf{y}$ where the SOI bears quaternary phase shift keying (QPSK) symbols using a root-raised cosine pulse-shaping filter with roll-off factor 0.5, spanning 8 QPSK symbols, and with an oversampling factor 16. The interference is an orthogonal frequency-division multiplexing (OFDM) signal. We generate an OFDM signal with symbols of length 80, bearing 16–quadrature amplitude modulation (QAM) symbols, with a fast Fourier transform (FFT) size of 64, and a cyclic-prefix of length 16. Details on the signals generation process are provided in the Github repository.[5]

A. The Potential Gain of Explicit Synchronization with DNNs

We now compare the performance of the DNN approach illustrated in Fig. 3 with the performance achieved by classical methods for detection and interference rejection, i.e., MF and the LMMSE estimator $\widehat{\mathbf{s}}_{\text{LMMSE}}$ given in (5), and by our proposed "synchronized" QLMMSE estimator $\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}}$ given in (15). For the CNN-based synchronization-to-interference methods

[4]To separate the communication signals used in this paper, other DNNs were implemented, yielding worse performance. Details can be found in the Github repository: https://github.com/RFChallenge/SCSS_DNN_Comparison.

[5]https://github.com/RFChallenge/SCSS_Sync

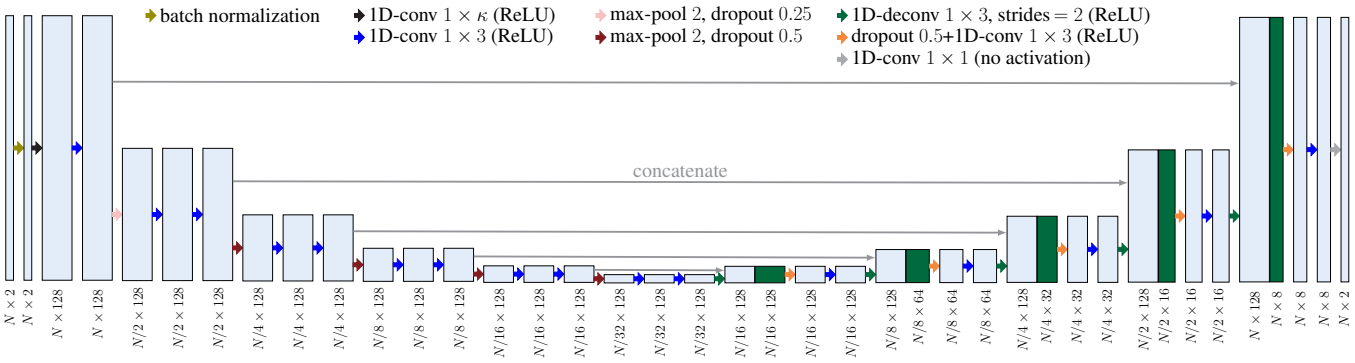

Fig. 4: Architecture of the DNN (U-Net) proposed to perform SCSS of the communication signals. The parameter κ denotes the kernel size of the first layer.

(Section III-C), we input 640 samples of the mixture $\mathbf{y}$ to the CNN. The input size to the separation U-Net is $N = 10240$.

In Fig. 5, we compare the performance in terms of BER as a function of the SIR in a noiseless setting. Specifically, we depict in gray the MF approach. In blue, we depict the LMMSE ($\widehat{\mathbf{s}}_{\text{LMMSE}}$ in (5)) computed using blocks of length 320.[6] In red, we depict the CNN–QLMMSE approach ($\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}}$ in (15)), also using blocks of length 320. Here, we explicitly synchronize to the interference signal, and exploit this to obtain "aligned statistics" (8) for each possible time-shift k_b. In green, we depict the performance of the U-Net approach when there is no explicit synchronization, i.e., the "Synchronization CNN" block in Fig. 3 is removed. Finally, we depict in black the DNN approach including both the synchronization and separation blocks, as described in Fig. 3, denoted as CNN–U-Net. Every described approach includes a last MF step before hard decoding based on the minimum Euclidean distance rule.

As can be observed, by only applying a MF to the received signal $\mathbf{y}$, which is optimal under white Gaussian noise, we do not exploit any temporal structure of the (non-Gaussian) interference. Hence, as expected, we obtain the worst performance. It is also evident that the LMMSE approach—optimal for Gaussian signals—without explicit alignment of the signal statistics via synchronization, is unable to exploit the underlying temporal nonstationarities, and accordingly yields approximately the performance obtained by only applying a MF to the received signal $\mathbf{y}$. However, by explicitly synchronizing to the interference signal using the CNN described in Section III-C, we can now use the *conditional* covariance of the interference for each possible time-shift k_b to obtain $\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}}$, which already leads to a significant performance gain. For example, for a BER of 10^{-3}, the CNN–QLMMSE approach requires an SIR of $-6\,\text{dB}$, while the MF and the LMMSE approaches require $-4\,\text{dB}$. Even though by explicitly synchronizing to the interference we can obtain significant gains, we recall that by using (quasi-)linear processing we can only exploit up to (conditional) second order statistics.

Since we consider digital communication signals, further gains can be achieved by exploiting high-order statistics

[6]For non-stationary processes, the required inversion of $\mathbf{C}_{yy}$ is computationally impractical for large N, as it is generally of complexity $\mathcal{O}(N^3)$.

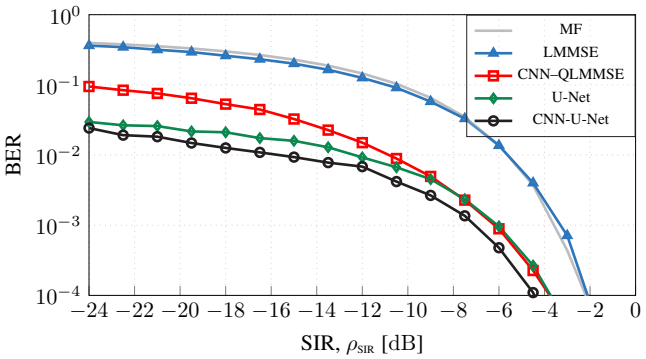

Fig. 5: BER as a function of the SIR for MF detection; LMMSE and QLMMSE interference rejection (blocks of length 320); and the data-driven U-Net approach with and without synchronization to the interference.

and the "discrete nature" of these signals. This is precisely achieved by our proposed DNN-based approaches (green and black). First, it is observed that a U-Net without prior explicit synchronization already outperforms the CNN–QLMMSE approach for most of the considered SIR values. The performance of the U-Net is further improved with explicit synchronization, using the block described in Fig. 1, as shown in Fig. 3. In this case, a BER of 10^{-2} is obtained at an SIR level of $-17\,\text{dB}$, while the U-Net without explicit synchronization requires $-12\,\text{dB}$, and the CNN–QLMMSE approach requires $-10.5\,\text{dB}$. Thus, for a given architecture with limited capacity (parametrization power), decoupling synchronization and separation can lead to considerable gains, which enables reliable communication in the presence of strong interference.

B. Gains from Explicit-Synchronization-Free Architecture

As mentioned in Section IV, a plausible interpretation of Lemma 1 is the following. When the input mixtures are sufficiently long, an explicit-synchronization-based architecture may not be required (or even provide superior performance), since the data is "very informative" with respect to the underlying time-shift. This essentially makes direct separation (i.e., an "implicit" synchronization approach) potentially preferable. Our best result up to date is achieved by directly inputting mixtures of length $N = 40960$ to the U-Net depicted in Fig. 4.

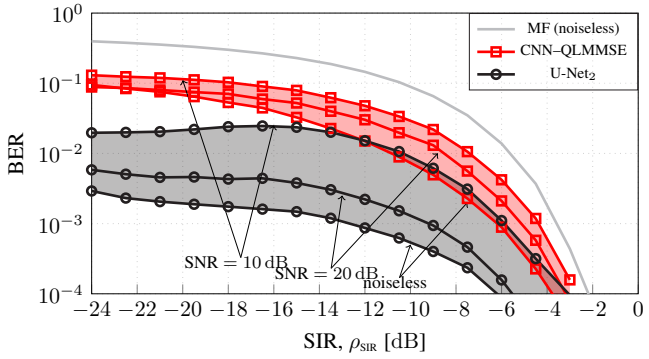

Fig. 6: BER as a function of the SIR for the noiseless case and SNR$=\{10, 20\}$ dB, for MF (noiseless only), QLMMSE interference rejection (blocks of length 320), and the CNN–U-Net approach described in Fig. 3.

Fig. 6 shows the performance of the U-Net scheme described in Fig. 4 (U-Net$_2$) where we input two replicas of the mixture $\mathbf{y}$, which provides the first layer with more diversity. We consider three different SNR levels of white Gaussian noise: $\rho_{\text{SNR}} \in \{10, 20, \infty\}$ dB. Specifically, we compare the performance of the DNN solution with the performance of the CNN–QLMMSE approach (computed using blocks of length 320) and the MF approach, which is only plotted for the noiseless case for the sake of clarity. Clearly, for all SNR levels, the U-Net$_2$ approach outperforms the CNN–QLMMSE and MF approaches. However, as expected, the smaller the SNR, the smaller the gap between them. For example, for a BER of 10^{-3}, the gain of the U-Net$_2$ approach compared to the CNN–QLMMSE is roughly 7 dB in the noiseless case, 4 dB for an SNR$=$ 20 dB, and 1.8 dB for an SNR$=$ 10 dB.

VI. Conclusions and Outlook

We study the SCSS problem with a focus on its application to interference rejection in digital communication. For Gaussian signals, we prove that a decoupled system architecture of synchronization followed by separation is asymptotically optimal in the MMSE sense. Consequently, since the optimal system can be impractical for implementation purposes, we propose a computationally attractive alternative with negligible performance loss relative to the optimal system. For (non-Gaussian) signals, we demonstrate in simulations that the proposed DNN-based data-driven approach can exploit the underlying temporal structures of the signals, thus leading to significant gains in terms of BER, and in particular, outperforms classical methods. Extensions of this work should focus on understanding how and when to use explicit synchronization in the context of SCSS with DNNs.

Appendix A
Proof of Lemma 1

To prove Lemma 1, we shall use the following lemma.

Lemma 2: For $\psi_N(\mathbf{y}, k)$ in Definition 1 (TDC), we have,

$$\mathbb{E}\left[e^{\tau\psi_N(\mathbf{y},\mathsf{k}_b)}\right] = \left(1-\frac{\tau}{N}\right)^N \cdot e^{-\tau}, \quad \forall \tau < N. \tag{16}$$

Proof of Lemma 2: First, recall $\mathbf{y}|\mathsf{k}_b \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{yy}(\mathsf{k}_b))$, where $\mathbf{C}_{yy}(\mathsf{k}_b) = \mathbf{C}_{ss}(0) + \mathbf{C}_{vv}(\mathsf{k}_b)$. Using the Cholesky decomposition, we write $\mathbf{C}_{yy}(\mathsf{k}_b) \triangleq \mathbf{\Gamma}_y(\mathsf{k}_b)\mathbf{\Gamma}_y^{\mathsf{H}}(\mathsf{k}_b)$, where $\mathbf{\Gamma}_y(\mathsf{k}_b) \in \mathbb{C}^{N\times N}$. Then, conditioned on k_b, we have

$$\psi_N(\mathbf{y}, \mathsf{k}_b) + 1 = \frac{1}{N}\Big(\underbrace{\mathbf{\Gamma}_y^{-1}(\mathsf{k}_b)\mathbf{y}}_{\triangleq \mathbf{u}(\mathsf{k}_b)}\Big)^{\mathsf{H}} \underbrace{\mathbf{\Gamma}_y^{-1}(\mathsf{k}_b)\mathbf{y}}_{=\mathbf{u}(\mathsf{k}_b)} = \frac{1}{N}\|\mathbf{u}(\mathsf{k}_b)\|_2^2, \tag{17}$$

where $\mathbf{u}(\mathsf{k}_b)|\mathsf{k}_b \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is a white Gaussian vector. By using the moment-generating function of $\frac{1}{N}\|\mathbf{u}(\mathsf{k}_b)\|_2^2$, which is a Gamma distributed random variable with shape parameter N and scale parameter $1/N$, we obtain

$$\mathbb{E}\left[\mathbb{E}\left[e^{\tau\left(\frac{1}{N}\|\mathbf{u}(\mathsf{k}_b)\|_2^2 - 1\right)}|\mathsf{k}_b\right]\right] \overset{\forall \tau < N}{=} \left(1-\frac{\tau}{N}\right)^{-N} e^{-\tau}. \tag{18}$$

■

By definition, the MAP estimator has the lowest error probability. Therefore, to show (13), it suffices to show that there exists another estimator of k_b, whose error probability is $o(N^{-\alpha})$ for any finite $\alpha \in \mathbb{R}_+$. To this end, consider the estimator,

$$\widehat{\mathsf{k}}_b \triangleq \arg\min_{m\in\mathcal{S}_{K_b}} |\psi_N(\mathbf{y}, m)|, \tag{19}$$

whose error probability is given by

$$\mathbb{P}\left[\widehat{\mathsf{k}}_b \neq \mathsf{k}_b\right] = \mathbb{P}\left[|\psi_N(\mathbf{y}, \mathsf{k}_b)| > \min_{m\in\mathcal{S}_{K_b}\setminus\mathsf{k}_b} |\psi_N(\mathbf{y}, m)|\right]. \tag{20}$$

We next show that, for some $0 < \epsilon < 0.5$,

$$\mathbb{P}\left[|\psi_N(\mathbf{y}, \mathsf{k}_b)| > N^{-0.5+\epsilon}\right] = o(N^{-\alpha}), \tag{21}$$

which, together with the the TDC in Definition 1, implies that $\mathbb{P}\left[\widehat{\mathsf{k}}_b \neq \mathsf{k}_b\right] = o(N^{-\alpha})$, and therefore (13). Note that

$$\lim_{N\to\infty} \psi_N(\mathbf{y}, \mathsf{k}_b) \overset{(\star)}{=} \mathbb{E}\left[\psi_N(\mathbf{y}, \mathsf{k}_b)\right] = \mathbb{E}\left[\mathbb{E}\left[\psi_N(\mathbf{y}, \mathsf{k}_b)|\mathsf{k}_b\right]\right] \tag{22}$$

$$= \frac{1}{N}\mathbb{E}\left[\mathbb{E}\left[\|\mathbf{u}(\mathsf{k}_b)\|_2^2|\mathsf{k}_b\right]\right] - 1 = 0, \tag{23}$$

where we have used (17), $\mathbf{u}(\mathsf{k}_b)|\mathsf{k}_b \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and $(\star)$ follows from the fact that $\mathbb{V}\mathrm{ar}(\psi_N(\mathbf{y}, \mathsf{k}_b)) = 1/N$, which can be shown in a similar fashion to (22)–(23).

Now, for any $a > 0$, it follows that

$$\mathbb{P}[|\psi_N(\mathbf{y}, \mathsf{k}_b)| > a] = \mathbb{P}[\psi_N(\mathbf{y}, \mathsf{k}_b) > a] + \mathbb{P}[\psi_N(\mathbf{y}, \mathsf{k}_b) < -a]. \tag{24}$$

Using the Chernoff bound and Lemma 2, we have, $\forall t < N$,

$$\mathbb{P}\left[\psi_N(\mathbf{y}, \mathsf{k}_b) > N^{-0.5+\epsilon}\right] \leq \left(1-\tfrac{t}{N}\right)^{-N} e^{-t\frac{N^{0.5-\epsilon}+1}{N^{0.5-\epsilon}}}, \tag{25}$$

$$\mathbb{P}\left[\psi_N(\mathbf{y}, \mathsf{k}_b) < -N^{-0.5+\epsilon}\right] \leq \left(1+\tfrac{t}{N}\right)^{-N} e^{-t\frac{N^{0.5-\epsilon}-1}{N^{0.5-\epsilon}}}. \tag{26}$$

Denote by $t_i^\star$, $i = \{1, 2\}$, the (easily obtained) minimizers of the right-hand side of (25) and (26) over t, respectively. Since for any $\alpha, \delta \in \mathbb{R}_+$ independent of N,

$$\lim_{N\to\infty} N^{\alpha+\delta}\left(1-\tfrac{t_1^\star}{N}\right)^{-N} e^{-t_1^\star\left(\frac{N^{0.5-\epsilon}+1}{N^{0.5-\epsilon}}\right)} = 0, \tag{27}$$

$$\lim_{N\to\infty} N^{\alpha+\delta}\left(1+\tfrac{t_2^\star}{N}\right)^{-N} e^{-t_2^\star\left(\frac{N^{0.5-\epsilon}-1}{N^{0.5-\epsilon}}\right)} = 0, \tag{28}$$

equation (21) follows, and this concludes the proof.

APPENDIX B
PROOF OF THEOREM 1

The proof follows by showing that, for any finite $\alpha \in \mathbb{R}_+$,

$$\varepsilon^2_{\text{MMSE}}(N) = \varepsilon^2_{\text{MAP-QLMMSE}}(N) + o\left(N^{-\alpha}\right). \tag{29}$$

Indeed, since $\varepsilon^2_{\text{MAP-QLMMSE}}(N) = \mathcal{O}(N)$, together with the trivial upper bound given by the definition of the MMSE estimator,

$$\varepsilon^2_{\text{MMSE}}(N) \leq \varepsilon^2_{\text{MAP-QLMMSE}}(N) \implies \frac{\varepsilon^2_{\text{MMSE}}(N)}{\varepsilon^2_{\text{MAP-QLMMSE}}(N)} \leq 1, \tag{30}$$

for any $N \in \mathbb{N}_+$, (29) implies (14).

We next show (29). By Lemma 1, α can always be chosen to obtain any polynomial rate. With slight abuse of notation, we write $o(N^{-\alpha})$ for the product of $o\left(N^{-\alpha}\right)$ with any polynomial in N of degree β (independent of N), which is $o(N^{-\alpha+\beta})$, and for a finite power, say $(\cdot)^\gamma$, of $o\left(N^{-\alpha}\right)$, which is $o\left(N^{-\alpha\gamma}\right)$.

By assumption, $\mathsf{k}_s = 0$. In this case, using the MAP-based QLMMSE (10), the MMSE estimator (7) can be expressed as

$$\begin{aligned}\widehat{\mathbf{s}}_{\text{MMSE}} = &\underbrace{\sum_{m_b \in \mathcal{S}_{K_b} \setminus \widehat{\mathsf{k}}_b^{\text{MAP}}} \mathbb{P}[\mathsf{k}_b = m_b|\mathbf{y}]\, \widehat{\mathbf{s}}_{\text{LMMSE}}(m_b)}_{\triangleq \boldsymbol{\delta}(\mathbf{y})} \\ &+ \mathbb{P}\left[\mathsf{k}_b = \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right] \widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{MAP}}).\end{aligned} \tag{31}$$

Using (31) and (10), we define the optimality gap (vector) as

$$\boldsymbol{\Delta}(\mathbf{y}) \triangleq \widehat{\mathbf{s}}_{\text{MMSE}} - \widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{MAP}}) \tag{32}$$

$$= \boldsymbol{\delta}(\mathbf{y}) - \mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right] \widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}. \tag{33}$$

Let $\mathbf{e}_{\text{MAP-QLMMSE}} \triangleq \widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} - \mathbf{s}$. It then follows that

$$\mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MMSE}} - \mathbf{s}\|_2^2\right] = \mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MMSE}} - \widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} + \widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} - \mathbf{s}\|_2^2\right] \tag{34}$$

$$= \varepsilon^2_{\text{MAP-QLMMSE}}(N) - \mathbb{E}\left[\|\boldsymbol{\Delta}(\mathbf{y})\|_2^2\right], \tag{35}$$

where we have used (32) in (34), and the well-known orthogonality property of the estimation error in MMSE estimation to any function of the measurements in (35). Hence, to prove (29), it is enough to show that $\mathbb{E}[\|\boldsymbol{\Delta}(\mathbf{y})\|_2^2] = o(N^{-\alpha})$. Expanding this term, we have,

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{\Delta}(\mathbf{y})\|_2^2\right] = &\,\mathbb{E}\left[\|\boldsymbol{\delta}(\mathbf{y})\|_2^2\right] \\ &- 2\Re\left\{\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right] \boldsymbol{\delta}^{\mathsf{H}}(\mathbf{y}) \widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\right]\right\} \\ &+ \mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]^2 \|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\|_2^2\right].\end{aligned} \tag{36}$$

We now show that the magnitude of each of the terms in (36) is bounded by $o(N^{-\alpha})$ terms. The first term in (36) reads

$$\begin{aligned}&\mathbb{E}\left[\|\boldsymbol{\delta}(\mathbf{y})\|_2^2\right] \\ &= \sum_{n=1}^{N} \mathbb{E}\left[\left(\sum_{m_b \in \mathcal{S}_{K_b} \setminus \widehat{\mathsf{k}}_b^{\text{MAP}}} \mathbb{P}[\mathsf{k}_b = m_b|\mathbf{y}]\, \widehat{\mathsf{s}}_{\text{LMMSE}}(m_b)\right)^2\right].\end{aligned} \tag{37}$$

By expanding the squared, using the Cauchy-Schwarz inequality twice, using that

$$\mathbb{P}[\mathbf{z} = \mathbf{z}]^\beta \leq \mathbb{P}[\mathbf{z} = \mathbf{z}], \quad \forall \beta \geq 1, \tag{38}$$

using Lemma 1, and that $\mathbb{E}\left[\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_1)\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_2)\right] = \mathcal{O}(1)$, it follows that $\mathbb{E}\left[\|\boldsymbol{\delta}(\mathbf{y})\|_2^2\right] = o(N^{-\alpha})$. The proof is concluded by applying similar steps to the second and third terms in (36)—Cauchy-Schwarz inequality together with (38), Lemma 1, $\mathbb{E}[(\boldsymbol{\delta}^{\mathsf{H}}(\mathbf{y})\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}})^2]^{\frac{1}{2}} = \mathcal{O}(N)$ (second term in (36)), and $\mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\|_2^4\right]^{\frac{1}{2}} = \mathcal{O}(N)$ (third term in (36))—such that each term is bounded by $o(N^{-\alpha})$.

REFERENCES

[1] 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.

[2] 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. 153 027–153 056, Aug. 2020.

[3] 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, Jun. 2016.

[4] Y. Gandelsman, A. Shocher, and M. Irani, ""Double-DIP": Unsupervised image decomposition via coupled deep-image-priors," in *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11 026–11 035.

[5] 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.

[6] M. G. Amin, "Interference mitigation in spread spectrum communication systems using time-frequency distributions," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 90–101, Jan. 1997.

[7] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[8] 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, Jul. 2019.

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

[10] 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.

[11] T. J. O'shea and N. West, "Radio machine learning dataset generation with GNU radio," in *Proc. GNU Radio Conf.*, vol. 1, no. 1, 2016.

[12] MIT RLE LIDS, "RF Challenge - AI Accelerator," accessed 2022-08-18. [Online]. Available: https://rfchallenge.mit.edu

[13] A. Napolitano, "Cyclostationarity: New trends and applications," *Signal Process.*, vol. 120, pp. 385–408, Mar. 2016.

[14] DeepSig Inc., "RF Datasets For Machine Learning," accessed 2022-08-18. [Online]. Available: https://www.deepsig.ai/datasets

[15] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 430–441, Mar. 2000.

[16] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. John Wiley & Sons, 2004.

[17] Z. Gao, C. Zhang, and Z. Wang, "Robust preamble design for synchronization, signaling transmission, and channel estimation," *IEEE Trans. Broadcast.*, vol. 61, no. 1, pp. 98–104, Jan. 2015.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer International Publishing, 2015, pp. 234–241.