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

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

***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) and synchronizing to both the SOI and the interference from a signal dataset enables 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 experiments. 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. Consequently, spectrum sharing will be unavoidable [1], [2]. Thus, different wireless communication systems will have to coexist in the same frequency bands, thereby generating unintentional interference 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.

Machine learning techniques has been successfully applied for *source separation* in the image and audio domains (see, e.g., [3], [4]. The solutions usually exploit the specific signals structure. For example, natural images may be separable by color features and local dependencies [5], while audio signals are usually separated by time-frequency spectrogram masking [6].

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. *These authors contributed equally to this work.

For communication signals, if the sources are separable in time and/or frequency, one could separate them via appropriate masking and classical filtering methods (see, e.g., [7]). The key challenge 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 [8], [9], 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 [10] and per-surviving processing algorithms [11] can be used. However, such methods require prior knowledge of the signal models, which in practice may not be known or available.

In such cases, 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., [12]). Hence, in this work, our goal is to capture information about the signals in a *data-driven* manner that can be exploited for SCSS. A natural choice is to use deep neural networks (DNNs). This problem has been recently promoted by the "RF Challenge" [13], where raw datasets of several radio-frequency signals are provided with minimal to no information about the signal generation processes.

In this paper, we study the data-driven SCSS problem when 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 (as is the case in many standard protocols), which are unknown *a priori*.[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 translates to better separation conditions. In our simulation results, we show that our proposed data-driven DNN approach is

[1]We assume the cyclic period is known. In practice, provided a dataset of the respective signal, this parameter can be consistently estimated [14].

superior to classical demodulation methods such as matched filtering (MF) applied to the received signal, and to classical interference rejection methods such as linear minimum mean-square error (LMMSE) estimation of the SOI. To improve upon these classical methods, designing the DNN according to the available domain knowledge is essential. We also prove that performing explicit consistent synchronization to the interference prior to separation, although suboptimal, can lead to asymptotically optimal separation, as the length of the received signal grows. This implies that, although a rich enough DNN performing joint synchronization and separation should outperform a two-step approach, for a given architecture, splitting synchronization and separation can lead to better performance, since the load of the DNN is now fully dedicated to separation, rather than to both synchronization and separation.

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 (components), given by

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

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 the signal $\mathsf{b}[n]$ as interference. The parameters $\mathsf{k}_s, \mathsf{k}_b \in \mathbb{Z}$ denote arbitrary (discrete) time-shifts with respect to the start of the cyclic periods of $\mathsf{s}[n]$ and $\mathsf{b}[n]$, respectively, which 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 signal component 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 value of $\mathbb{E}[\ell(\widehat{\mathbf{s}}, \mathbf{s}(\mathsf{k}_s))]$ is minimized. This problem is referred to as SCSS.

In this work, we assume we do not have precise knowledge of the underlying distributions of the SOI and the interference. However, we assume the avaiability of a dataset wherein the signals and their respective time-shifts $(\mathbf{s}(k_s), k_s)$ and $(\mathbf{b}(k_b), k_b)$ are known, which allows for a data-driven approach to solve the problem. Examples of these kind of datasets can be found in [13], [15].

III. The Gain in Synchronization to Interference

Before we present our approach to the SCSS problem introduced in Section II, we provide an analysis of an asymptotically optimal estimator of $\mathbf{s}(\mathsf{k}_s)$ (for a given reconstruction metric ℓ), which will shed light on key aspects in optimal separation and the role of *synchronization to interference*.

In the following analysis, we focus on minimum mean-square error (MMSE) optimality, namely, we choose $\ell(\mathbf{x}, \mathbf{z}) \triangleq \|\mathbf{x} - \mathbf{z}\|_2^2$, and we assume that $\mathsf{s}[n]$ and $\mathsf{b}[n]$ are Gaussian processes, which is a widely accepted assumption to model communication signals [16]. In this case, we define[2]

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

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

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

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 (in the form of $\mathbf{C}_{vv}$) that can be exploited for enhanced separation performance.

A. Linear 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 so-called LMMSE estimator [17], given by (assuming $\det(\mathbf{C}_{yy}) \neq 0$)

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

is constructed using the statistics of the mixture that inherently take into account the potentially non-trivial structure of $\mathbf{C}_{vv}$ i.e., a 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 restriction. For example, 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 individually 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.

[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.

B. MMSE Estimation

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

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

whose mean-square 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 form

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

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 we denote $\widehat{\mathbf{s}}_{\text{LMMSE}}(m_s,m_b) \triangleq \left[\mathbf{C}_{ss}(m_s)+\mathbf{C}_{vv}(m_b)\right]^{-1}\mathbf{C}_{ss}(m_s)\mathbf{y}$, with $\mathbf{C}_{ss}(m) \triangleq \mathbb{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}|\mathsf{k}_s=m]$ and $\mathbf{C}_{vv}(m) \triangleq \mathbb{E}[\mathbf{v}\mathbf{v}^{\mathsf{H}}|\mathsf{k}_b=m]$. Put simply, (7) is a weighted average of $K_s\cdot K_b$ linear estimators, with the posterior probabilities—which are *nonlinear* functions of the data $\mathbf{y}$—serving as the weights. Even before taking into account the computation of the posteriors, the sum in (7) scales with the number of possible time-shifts K_s and K_b, rendering $\widehat{\mathbf{s}}_{\text{MMSE}}$ often impractical.

As can be seen from (7), synchronization, namely, perfect knowledge of the time-shifts, already substantially simplifies the computation, since in that case all that is required is the (conditional) linear estimator $\widehat{\mathbf{s}}_{\text{LMMSE}}(m_s,m_b)$. In other words, eliminating this type of randomness from the mixture $\mathbf{y}$ grants us the lower complexity and simple form of a linear estimator. Fortunately, a two-step "synchronization-separation" estimator can approach the MMSE, 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 signal of interest,[3] namely $\mathsf{k}_s=0$. However, the result below can be generalized to the case where the SOI's time-shift k_s is random and unknown. For a concise presentation of the theorem, we use the following quantities.

Let $\widehat{\mathsf{k}}_b^{\text{MAP}} \triangleq \arg\max_{m\in\mathcal{S}_{K_b}} \mathbb{P}[\mathsf{k}_b=m|\mathbf{y}]$ be the maximum *a posteriori* (MAP) estimator of k_s, and define the (suboptimal) "plug-in", MAP-based *quasi*-linear MMSE estimator

$$\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} \triangleq \widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{MAP}}) \in \mathbb{C}^{N\times 1}, \tag{8}$$

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

$$\varepsilon^2_{\text{MMSE}}(N) \triangleq \mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MMSE}}-\mathbf{s}\|_2^2\right] \in \mathbb{R}_+, \tag{9}$$

$$\varepsilon^2_{\text{MAP-QLMMSE}}(N) \triangleq \mathbb{E}[\|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}-\mathbf{s}\|_2^2] \in \mathbb{R}_+. \tag{10}$$

Definition 1 ("temporal-diversity" condition (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 $k\in\mathcal{S}_{K_b}/\mathsf{k}_b$ such that $\lim_{N\to\infty}|\psi_N(\mathbf{y},k)|=0$.

[3]This is a reasonable assumption in most communication systems [18].

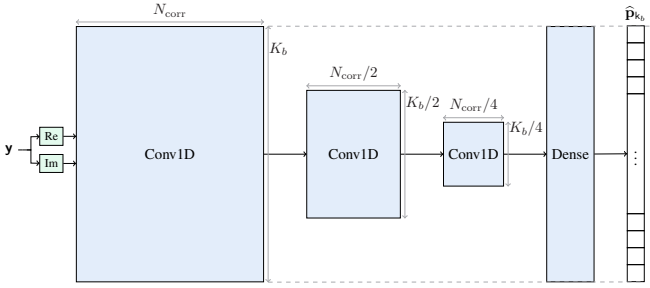

Fig. 1: Architecture of the proposed CNN for synchronization.

Theorem 1: Under the TDC, we have

$$\lim_{N\to\infty}\frac{\varepsilon^2_{\text{MMSE}}(N)}{\varepsilon^2_{\text{MAP-QLMMSE}}(N)}=1. \tag{11}$$

Proof: See Appendix A. ∎

In words, Theorem 1 says that first successfully 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 [19, Fig. 1].

C. Synchronization via convolutional neural networks (CNNs)

Although the estimator (8) 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 $\widehat{\mathsf{k}}_s^{\text{MAP}}$. However, when by measurement or generation, sufficiently large datasets of 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 to capture 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, $\mathcal{D}_T \triangleq \{(\mathbf{y}^{(i)},\mathsf{k}_b^{(i)}) : i\in\mathcal{S}_{I_T}\}$, 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 to the strongest, most informative (for estimation) temporal structures. Since the cyclic period K_b is assumed to be known, we train a model using the the crossentropy 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 the estimator $\widehat{\mathsf{k}}_b^{\text{CNN}} \triangleq \arg\max_{m\in\mathcal{S}_{K_b}} \mathbf{e}_m^{\mathsf{T}}\widehat{\mathbf{p}}_{\mathsf{k}_b}$, cf. $\widehat{\mathsf{k}}_b^{\text{MAP}}$, where $\mathbf{e}_m\in\mathbb{R}^{K_b\times 1}$ denotes the standard basis vector.

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

$$\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}} \triangleq \widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{CNN}}), \tag{12}$$

for the same communication waveforms described in detail in Section V, but considering here Gaussian alphabets instead of

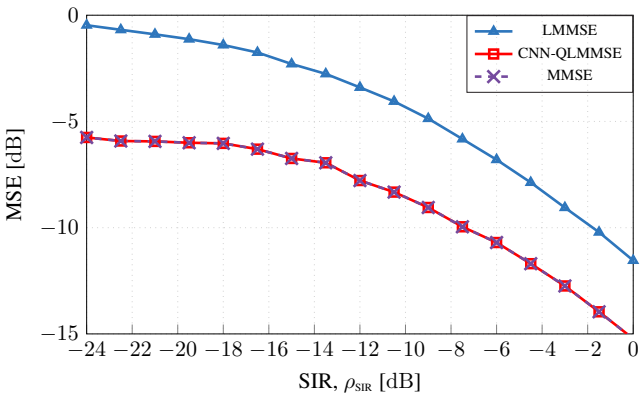

Fig. 2: MSE as a function of the SIR (ρ_{SIR}) for a fixed SNR (ρ_{SNR}) of 20 dB.

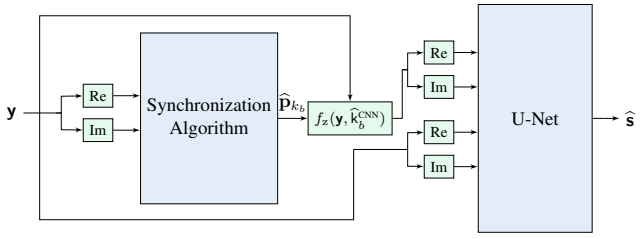

Fig. 3: System architecture of the DNN–based approach with an explicit CNN-based synchronization block prior to the separation block (U-Net).

the discrete and finite alphabets considered 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 (10), which asymptotically coincides with the MMSE (9), by virtue of Theroem 1.

Thus, 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 next present a supervised learning approach for SCSS, which is used in this work as an interference rejection method.

The DNN architecture proposed for SCSS is depicted in Fig. 3. As can be observed, the architecture consists of two main building blocks: (i) CNN to perform synchronization to 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, which is suboptimal, can lead to optimal separation asymptotically as N grows. Second, even though a rich enough DNN might be able to perform implicitly the synchronization task jointly with the separation, for a given architecture, acquiring this knowledge explicitly helps to reduce the complexity of the separation task. In Section V, we show that this decoupled approach provides substantial performance gains.

The synchronization block is based on the CNN described in Section III-C (see also Fig. 1). The DNN for separation is based on the so-called U-Net (see Fig. 4), which was first proposed for biomedical image segmentation in [20], and has some specific properties that makes it suitable to the specific informative features of digital communication signals. In particular, its CNN building blocks allow us to input long signals (e.g., $N = 40960$) that allows for processing of long time intervals, which cannot be processed using classical methods. In turn, processing such long observations allows for exploitation of temporal structures on a different scale, which naturally leads to performance gains.

As shown in Fig. 4, our DNNs approach departs from standard implementations intended to deal with images, which are 2D signals. To handle 1D complex-valued, time-series communication signals, we use 1D convolutional layers. Furthermore, differently from standard CNN-based architectures designed originally to deal with images that uses 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, to ensure capturing the most relevant temporal structure components of both the SOI and interference. This results in one order of magnitude gains.

For training, we stack the real and imaginary parts of the signals as separate channels to both the synchronization-to-interference CNN and the separation U-Net. Specifically, for separation, we input the stacked real and imaginary parts of the input $\mathbf{y}$, and the stacked real and imaginary parts of the output of $f_z(\mathbf{y}, \widehat{\mathsf{k}}_b)$, which simply implements zero-padding on the first $\widehat{\mathsf{k}}_b$ symbols of $\mathbf{y}$, as a means to provide the additional information about $\widehat{\mathsf{k}}_b$ to the U-Net. This implicitly imposes some regularization according to the estimated time-shift $\widehat{\mathsf{k}}_b$, obtained during the synchronization-to-interference stage. 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 the Github repository.[5]

V. Numerical Results

We consider mixtures $\mathbf{y}$ of length $N = 40960$, where the SOI bears quaternary phase shift keying (QPSK) symbols using a root-raised cosine pulse-shaping filter with roll-off factor equal to 0.5, spanning 8 QPSK symbols, and with over-sampling factor equal to 16. The interference is an orthogonal frequency-division multiplexing (OFDM) signal. Specifically, we consider 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. Further details on the signals generation process are provided in the Github repository.[5]

We next compare the performance of the DNN approach illustrated in Fig. 3 (described in Sections III-C and IV) with

[4]For the communication signals used in this paper, other DNNs were implemented, yielding worse performance. Full details can be found in the Github repository: copyurl.

[5]Github repository: copy url here

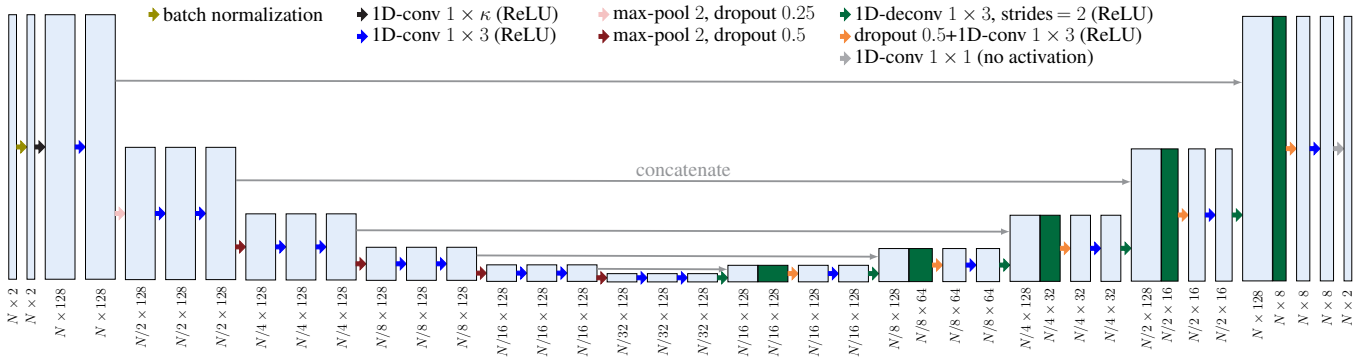

Fig. 4: Architecture of the DNN (U-Net) proposed to perform SCSS of the communication signals considered in this paper.

the performance achieved by classical methods for detection and interference rejection, i.e., MF to the SOI at the receiver; the classical LMMSE estimator $\widehat{\mathbf{s}}_{\text{LMMSE}}$ given in (5); and our propsed "synchronized" QLMMSE estimator $\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}}$ given in (12). For the cases including synchronization to interference via the CNN described in Section III-C, we input 640 samples from the mixture $\mathbf{y}$ to such a DNN. Since the input size to the separation block is $N = 40960$, we highlight that the synchronization block uses *only* the first $N/64$ samples of each mixture example, hence is relatively computationally efficient. In the U-net, we fix the long kernel size $\kappa = 101$.

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 (12)), also using blocks of length 320. Here, we explicitly synchronize to the interference signal, and exploit this to obtain aligned statistics for each possible time-shift k_b. In green, we depict the performance of the U-Net approach, where there is no explicit synchronization, hence the inputs coming from the synchronization block are not used. Finally, we depict in black the DNN approach including both the synchronization and separation blocks, as described in Fig. 3, which we denote 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 interference, we do not exploit any temporal structure of the (non-Gaussian) interference, hence 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 compute the *conditional* covariance

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

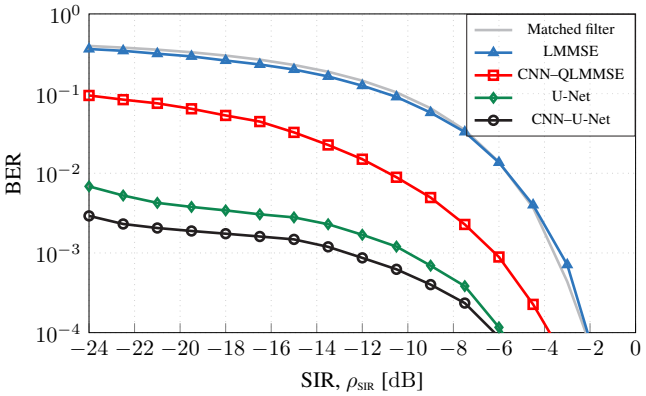

Fig. 5: BER as a function of the SIR for MF detection; LMMSE and QLMMSE interference rejection (blocks of length 320); and the data-driven U-Net approach with and without synchronization to the interference.

of the interference for each possible time-shift k_b to obtain $\widehat{\mathbf{s}}_{\text{CNN-QLMMSE}}$, which 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. By the nature of the considered communication signals, further gains can be achieved by exploiting higher order statistics and the "discrete nature" of the signals. This is precisely achieved by our proposed DNN-based approaches (red and black). First, it is observed that a U-Net without prior explicit synchronization clearly outperforms the CNN–QLMMSE approach. For example, it achieves a BER below 10^{-3} at an SIR level of $-10\,\text{dB}$, while the CNN–QLMMSE approach requires $-6\,\text{dB}$. Finally, the performance of the U-Net are further improved with explicit synchronization, using the block as described in Fig. 1. In this case, a BER of 10^{-3} is already obtained at an SIR level of $-12\,\text{dB}$, while the U-Net without explicit synchronization requires $-10\,\text{dB}$. Thus, for a given architecture with limited capacity (parametrization power), decoupling synchronization and separation can lead to considerable gains.

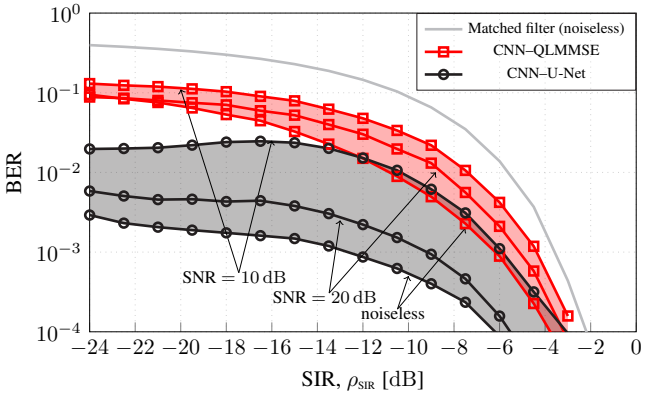

Fig. 6: BER as a function of the SIR for the noiseless case and SNR= $\{10, 20\}$ dB for MF (noiseless only), QLMMSE interference rejection (blocks of length 320), with the data-driven U-Net approach described in Fig. 3.

Finally, in Fig. 6 we show the performance of the DNN scheme described in Fig. 3 (CNN–U-Net) under the presence of white Gaussian noise, for three different SNR levels: $\{10, 20, \infty\}$ 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 CNN–U-Net approach outperforms the CNN–QLMMSE approach. 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 CNN–U-Net 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

We studied the SCSS problem with a focus on its application to interference rejection in the context of digital communication. For Gaussian signals, we proved that a decoupled system architecture of synchronization followed by separation is asymptotically optimal in the MMSE sense. Consequently, while the optimal system is impractical for implementation purposes, we proposed a computationally attractive alternative with negligible performance loss relative to the optimal system. For (non-Gaussian) communication signals, we demonstrated in simulations that the proposed DNN-based data-driven approach can exploit the underlying temporal structures of the signals by explicit synchronization, thus leading to significant gains in terms of BER, and in particular, outperforms classical methods. Extensions of this work include, among other, "soft" synchronization, and further refinements of the training loss and architectural choices.

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, 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, 2020.

[3] D. Stoller, S. Ewert, and S. Dixon, "Wave-U-Net: A multi-scale neural network for end-to-end audio source separation," *arXiv:1806.03185*, 2018.

[4] 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, 2016.

[5] 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)*, 2019, pp. 11 026–11 035.

[6] 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, 2015.

[7] 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, 1997.

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

[9] 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, 2019.

[10] T. Shilong, C. Shaohe, Z. Hui, and W. Jian, "Particle filtering based single-channel blind separation of co-frequency MPSK signals," in *2007 International Symposium on Intelligent Signal Processing and Communication Systems*, 2007, pp. 582–585.

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

[12] 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.

[13] MIT RLE, "RF Challenge - AI Accelerator," https://rfchallenge.mit.edu, accessed 2022-04-04. [Online]. Available: https://rfchallenge.mit.edu

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

[15] DeepSig Inc., "RF Datasets For Machine Learning," https://www.deepsig.ai/datasets, accessed 2022-04-04. [Online]. Available: https://www.deepsig.ai/datasets

[16] 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, 2000.

[17] H. L. Van Trees, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

[18] 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, 2015.

[19] 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, 2019.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

Appendix A
Proof of Theorem 1

To prove the theorem, we shall use the following lemma, whose proof is given in Appendix B.

Lemma 1: For any finite $\alpha \in \mathbb{R}_+$, we have

$$\mathbb{P}\left[\widehat{\mathsf{k}}_b^{\text{MAP}} \neq \mathsf{k}_b\right] = \mathcal{O}\left(\frac{1}{N^\alpha}\right). \tag{13}$$

From Lemma 1, we have the following corollary.

Corollary 1: Using (13), for any finite $\alpha \in \mathbb{R}_+$, we have

$$\mathbb{E}\left[\mathbb{P}\left[\widehat{\mathsf{k}}_b^{\text{MAP}} \neq \mathsf{k}_b|\mathbf{y}\right]\right] = \mathcal{O}\left(\frac{1}{N^\alpha}\right), \tag{14}$$

$$\mathbb{E}\left[\mathbb{P}\left[\widehat{\mathsf{k}}_b^{\text{MAP}} = k|\mathbf{y}\right]\right] = \mathcal{O}\left(\frac{1}{N^\alpha}\right), \ \forall k \in \mathcal{S}_{K_b}/\mathsf{k}_b. \tag{15}$$

The roadmap for the proof of the theorem is as follows:

- Step 1: Express the optimality gap between the MMSE and MAP-based QLMMSE estimators as a function of the error probability of the MAP synchronizer $\widehat{\mathsf{k}}_b^{\text{MAP}}$.
- Step 2: Upper and lower bound the MMSE (9) in terms of the MAP-based QLMMSE (10) and the optimality gap.
- Step 3: Show that the upper and lower bounds are either equal to or approach (10) as $N \to \infty$ (sufficiently fast), and apply the squeeze theorem to the ratio in (11).

Proof of Theorem 1: Let us write the the MMSE estimator (6), explicitly, using (7), in terms of the MAP-based QLMMSE estimator (8), as (recall $\mathsf{k}_s = 0$, by assumption),

$$\begin{aligned}\widehat{\mathbf{s}}_{\text{MMSE}} &= \sum_{m_b=1}^{K_b} \mathbb{P}[\mathsf{k}_b = m_b|\mathbf{y}]\,\widehat{\mathbf{s}}_{\text{LMMSE}}(m_b) \\ &= \underbrace{\sum_{\substack{m_b=1 \\ m_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \mathbb{P}[\mathsf{k}_b = m_b|\mathbf{y}]\,\widehat{\mathbf{s}}_{\text{LMMSE}}(m_b)}_{\triangleq \boldsymbol{\delta}(\mathbf{y})} \\ &\quad + \mathbb{P}\left[\mathsf{k}_b = \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]\widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{MAP}}).\end{aligned} \tag{16}$$

Using (16), we define the optimality gap (vector),

$$\boldsymbol{\Delta}(\mathbf{y}) \triangleq \widehat{\mathbf{s}}_{\text{MMSE}} - \widehat{\mathbf{s}}_{\text{LMMSE}}(\widehat{\mathsf{k}}_b^{\text{MAP}}) = \widehat{\mathbf{s}}_{\text{MMSE}} - \widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} \tag{17}$$

$$= \boldsymbol{\delta}(\mathbf{y}) - \mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}. \tag{18}$$

At this point, our strategy for the proof is to "sandwich" the ratio in the limit (11) by upper and lower bounds that are equal or converge to 1. The upper bound is trivially given by the definition of the MMSE estimator, since, for any $N \in \mathbb{N}_+$,

$$\varepsilon^2_{\text{MMSE}}(N) \leq \varepsilon^2_{\text{MAP-QLMMSE}}(N) \implies \frac{\varepsilon^2_{\text{MMSE}}(N)}{\varepsilon^2_{\text{MAP-QLMMSE}}(N)} \leq 1. \tag{19}$$

We proceed to the lower bound. For shorthand, let $\mathbf{e}_{\text{MAP-QLMMSE}} \triangleq \widehat{\mathbf{s}}_{\text{MAP-QLMMSE}} - \mathbf{s}$, and let us first write the MMSE in terms of the estimation error $\mathbf{e}_{\text{MAP-QLMMSE}}$ and the optimality gap $\boldsymbol{\Delta}(\mathbf{y})$, and further lower bound it, as,

$$\mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MMSE}} - \mathbf{s}\|_2^2\right] = \tag{20}$$

$$\mathbb{E}\left[\|\mathbf{e}_{\text{MAP-QLMMSE}} + \boldsymbol{\Delta}(\mathbf{y})\|_2^2\right] \geq \tag{21}$$

$$\mathbb{E}\left[\left(\|\mathbf{e}_{\text{MAP-QLMMSE}}\|_2 - \|\boldsymbol{\Delta}(\mathbf{y})\|_2\right)^2\right] = \tag{22}$$

$$\varepsilon^2_{\text{MAP-QLMMSE}}(N) + \mathbb{E}\left[\|\boldsymbol{\Delta}(\mathbf{y})\|_2^2\right] - 2\,\mathbb{E}\left[\|\mathbf{e}_{\text{MAP-QLMMSE}}\|_2\|\boldsymbol{\Delta}(\mathbf{y})\|_2\right], \tag{23}$$

where we have used (17) in (20), $\|\mathbf{x}-\mathbf{y}\| \geq |\|\mathbf{x}\| - \|\mathbf{y}\||$ in (21), and (10) in (22). We now focus on the last two terms in (23). Starting with the term in the middle, we have,

$$\begin{aligned}&\mathbb{E}\left[\|\boldsymbol{\Delta}(\mathbf{y})\|_2^2\right] = \\ &\mathbb{E}\left[\|\boldsymbol{\delta}(\mathbf{y})\|_2^2\right] + \mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]^2 \|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\|_2^2\right] \\ &\qquad + 2\Re\left\{\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]\boldsymbol{\delta}^{\mathsf{H}}(\mathbf{y})\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\right]\right\}.\end{aligned} \tag{24}$$

We now show that (the magnitude of) each of the terms in (24) is bounded. It will then follow that the expected squared norm of the optimality gap, $\mathbb{E}\left[\|\boldsymbol{\Delta}(\mathbf{y})\|_2^2\right]$, is also bounded.

Starting with the first term in (24), we have,

$$\mathbb{E}\left[\|\boldsymbol{\delta}(\mathbf{y})\|_2^2\right] = \sum_{n=1}^{N} \mathbb{E}\left[\delta_n^2(\mathbf{y})\right] = \tag{25}$$

$$\sum_{n=1}^{N} \mathbb{E}\left[\left(\sum_{\substack{m_b=1 \\ m_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \mathbb{P}[\mathsf{k}_b = m_b|\mathbf{y}]\,\widehat{\mathsf{s}}_{\text{LMMSE},n}(m_b)\right)^2\right]. \tag{26}$$

Focusing on one element of the sum in (26), we have,

$$\mathbb{E}\left[\left(\sum_{\substack{m_b=1 \\ m_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \mathbb{P}[\mathsf{k}_b = m_b|\mathbf{y}]\,\widehat{\mathsf{s}}_{\text{LMMSE},n}(m_b)\right)^2\right] \leq \tag{27}$$

$$\sum_{\substack{m_1=1 \\ m_1 \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \sum_{\substack{m_2=1 \\ m_2 \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \mathbb{E}\left[\mathbb{P}[\mathsf{k}_b = m_1|\mathbf{y}]^2\,\mathbb{P}[\mathsf{k}_b = m_2|\mathbf{y}]^2\right]^{\frac{1}{2}} \cdot$$

$$\mathbb{E}\left[\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_1)\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_2)\right]^{\frac{1}{2}} \leq \tag{28}$$

$$\sum_{\substack{m_1=1 \\ m_1 \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \sum_{\substack{m_2=1 \\ m_2 \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \mathbb{E}\left[\mathbb{P}[\mathsf{k}_b = m_1|\mathbf{y}]^4\right]^{\frac{1}{4}} \mathbb{E}\left[\mathbb{P}[\mathsf{k}_b = m_2|\mathbf{y}]^4\right]^{\frac{1}{4}} \cdot$$

$$\mathbb{E}\left[\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_1)\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_2)\right] \leq \tag{29}$$

$$\sum_{\substack{m_1=1 \\ m_1 \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \sum_{\substack{m_2=1 \\ m_2 \neq \widehat{\mathsf{k}}_b^{\text{MAP}}}}^{K_b} \underbrace{\mathbb{E}\left[\mathbb{P}[\mathsf{k}_b = m_1|\mathbf{y}]\right]^{\frac{1}{4}}}_{\mathcal{O}\left(\frac{1}{N^{4\alpha}}\right)} \underbrace{\mathbb{E}\left[\mathbb{P}[\mathsf{k}_b = m_2|\mathbf{y}]\right]^{\frac{1}{4}}}_{\mathcal{O}\left(\frac{1}{N^{4\alpha}}\right)} \cdot \tag{30}$$

$$\underbrace{\mathbb{E}\left[\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_1)\widehat{\mathsf{s}}^2_{\text{LMMSE},n}(m_2)\right]}_{\mathcal{O}(1)} = \tag{31}$$

$$\mathcal{O}\left(\frac{1}{N^\alpha}\right), \tag{32}$$

where we have used the Cauchy-Schwarz inequality repeatedly in (27) and (28), the following, almost trivial, observation,

$$\mathbb{P}[\mathbf{z} = \mathbf{z}]^\beta \leq \mathbb{P}[\mathbf{z} = \mathbf{z}], \quad \forall \beta \geq 1. \tag{33}$$

in (29), and (15) in (30). Since (26) is a sum of N terms as in (27), we obtain

$$\mathbb{E}\left[\|\boldsymbol{\delta}(\mathbf{y})\|_2^2\right] = \mathcal{O}\left(\frac{1}{N^{\alpha-1}}\right). \tag{34}$$

Moving to the second term in (24), we have,

$$\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]^2 \|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\|_2^2\right] \leq \tag{35}$$

$$\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right] \|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\|_2^2\right] \leq \tag{36}$$

$$\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]^2\right]^{\frac{1}{2}} \mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\|_2^4\right]^{\frac{1}{2}} \leq \tag{37}$$

$$\underbrace{\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]\right]^{\frac{1}{2}}}_{\mathcal{O}\left(\frac{1}{N^{2\alpha}}\right)} \underbrace{\mathbb{E}\left[\|\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\|_2^4\right]^{\frac{1}{2}}}_{\mathcal{O}(N)} = \mathcal{O}\left(\frac{1}{N^{\alpha-1}}\right), \tag{38}$$

where we have used (33) in (35) and (37), the Cauchy-Schwarz inequality in (36), and (14) in (38). As for the magnitude of the last term in (24), we similarly obtain,

$$\left|\Re\left\{\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]\boldsymbol{\delta}^{\mathsf{H}}(\mathbf{y})\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\right]\right\}\right| \leq \tag{39}$$

$$\left|\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]\boldsymbol{\delta}^{\mathsf{H}}(\mathbf{y})\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\right]\right| \leq \tag{40}$$

$$\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]^2\right]^{\frac{1}{2}} \mathbb{E}\left[\left(\boldsymbol{\delta}^{\mathsf{H}}(\mathbf{y})\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\right)^2\right]^{\frac{1}{2}} \leq \tag{41}$$

$$\underbrace{\mathbb{E}\left[\mathbb{P}\left[\mathsf{k}_b \neq \widehat{\mathsf{k}}_b^{\text{MAP}}|\mathbf{y}\right]\right]^{\frac{1}{2}}}_{\mathcal{O}\left(\frac{1}{N^{2\alpha}}\right)} \underbrace{\mathbb{E}\left[\left(\boldsymbol{\delta}^{\mathsf{H}}(\mathbf{y})\widehat{\mathbf{s}}_{\text{MAP-QLMMSE}}\right)^2\right]^{\frac{1}{2}}}_{\mathcal{O}(N)} = \tag{42}$$

$$\mathcal{O}\left(\frac{1}{N^{\alpha-1}}\right), \tag{43}$$

where we have used, again, the Cauchy-Schwarz inequality in (40), (33) in (41), and (14) in (42). We note in passing that the term on the right in (42) may be bound more tightly, but this is not necessary for the following steps of this proof.

As an intermediate summary, we have established upper bounds on the magnitudes of the terms in (24). Hence, using (34), (38) and (43), we now have

$$\mathbb{E}\left[\|\boldsymbol{\Delta}(\mathbf{y})\|_2^2\right] = \mathcal{O}\left(\frac{1}{N^{\alpha-1}}\right). \tag{44}$$

Proceeding to the last term in (23), we have,

$$\mathbb{E}\left[\|\mathbf{e}_{\text{MAP-QLMMSE}}\|_2 \|\Delta(\mathbf{y})\|_2\right] \leq \tag{45}$$

$$\underbrace{\mathbb{E}\left[\|\mathbf{e}_{\text{MAP-QLMMSE}}\|_2\right]^{\frac{1}{2}}}_{\mathcal{O}(N)} \underbrace{\mathbb{E}\left[\|\Delta(\mathbf{y})\|_2\right]^{\frac{1}{2}}}_{\mathcal{O}\left(\frac{1}{N^{\alpha-1}}\right)} = \mathcal{O}\left(\frac{1}{N^{\frac{\alpha}{2}-1}}\right) \tag{46}$$

$$\implies \mathbb{E}\left[\|\mathbf{e}_{\text{MAP-QLMMSE}}\|_2 \|\Delta(\mathbf{y})\|_2\right] = \mathcal{O}\left(\frac{1}{N^{\frac{\alpha}{2}-1}}\right). \tag{47}$$

where we have used the Cauchy-Schwarz inequality in (45), and (34) in (46). Now, with (44) and (47) established above, revisiting the lower bound (23), we have,

$$\varepsilon^2_{\text{MMSE}}(N) \geq \varepsilon^2_{\text{MAP-QLMMSE}}(N) + \mathcal{O}\left(\frac{1}{N^{\frac{\alpha}{2}-1}}\right) \tag{48}$$

$$\implies \frac{\varepsilon^2_{\text{MMSE}}(N)}{\varepsilon^2_{\text{MAP-QLMMSE}}(N)} \geq 1 + \mathcal{O}\left(\frac{1}{N^{\frac{\alpha}{2}}}\right). \tag{49}$$

By combining the upper (19) and lower (49) bounds, choosing, for example (since (13) is for any positive α), $\alpha = 2$, and taking the limit $N \to \infty$, (11) is established. ∎

APPENDIX B
PROOF OF LEMMA 1

In the proof of Lemma 1, we shall use the following lemma.

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

$$\mathbb{E}\left[e^{\tau\psi_N(\mathbf{y},\mathsf{k}_b)}\right] = \left(1 - \frac{2\tau}{N}\right)^{-\frac{N}{2}} \cdot e^{-\tau}, \ \ \forall \tau < \frac{N}{2}. \tag{50}$$

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 $\mathbf{C}_{yy}(\mathsf{k}_b) \triangleq \boldsymbol{\Gamma}_y(\mathsf{k}_b)\boldsymbol{\Gamma}_y^{\mathsf{H}}(\mathsf{k}_b)$, where $\boldsymbol{\Gamma}_y(\mathsf{k}_b) \in \mathbb{C}^{N\times N}$. Then, conditioned on k_b, we have

$$\psi_N(\mathbf{y}, \mathsf{k}_b) + 1 = \frac{1}{N}\mathbf{y}^{\mathsf{H}}\mathbf{C}_{yy}^{-1}(\mathsf{k}_b)\mathbf{y} \tag{51}$$

$$= \frac{1}{N}\mathbf{y}^{\mathsf{H}}\boldsymbol{\Gamma}_y^{-\mathsf{H}}(\mathsf{k}_b)\boldsymbol{\Gamma}_y^{-1}(\mathsf{k}_b)\mathbf{y} \tag{52}$$

$$= \frac{1}{N}\underbrace{\left(\boldsymbol{\Gamma}_y^{-1}(\mathsf{k}_b)\mathbf{y}\right)}_{\triangleq\mathbf{u}(\mathsf{k}_b)}^{\mathsf{H}} \underbrace{\boldsymbol{\Gamma}_y^{-1}(\mathsf{k}_b)\mathbf{y}}_{=\mathbf{u}(\mathsf{k}_b)} \tag{53}$$

$$= \frac{1}{N}\|\mathbf{u}(\mathsf{k}_b)\|_2^2, \tag{54}$$

where $\mathbf{u}(\mathsf{k}_b)|\mathsf{k}_b \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is a white Gaussian vector. Thus,

$$\mathbb{E}\left[e^{\tau\psi_N(\mathbf{y},\mathsf{k}_b)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\tau\psi_N(\mathbf{y},\mathsf{k}_b)}|\mathsf{k}_b\right]\right] \tag{55}$$

$$= \mathbb{E}\left[\mathbb{E}\left[e^{\tau\left(\frac{1}{N}\|\mathbf{u}(\mathsf{k}_b)\|_2^2 - 1\right)}|\mathsf{k}_b\right]\right] \tag{56}$$

$$= \mathbb{E}\left[\mathbb{E}\left[e^{\frac{\tau}{N}\sum_{n=1}^N \mathsf{u}_n^2(\mathsf{k}_b)}|\mathsf{k}_b\right]\right] e^{-\tau} \tag{57}$$

$$= \mathbb{E}\left[\prod_{n=1}^N \mathbb{E}\left[e^{\frac{\tau}{N}\mathsf{u}_n^2(\mathsf{k}_b)}|\mathsf{k}_b\right]\right] e^{-\tau} \tag{58}$$

$$\overset{\forall\tau<\frac{N}{2}}{=} \mathbb{E}\left[\prod_{n=1}^N \left(1 - \frac{2\tau}{N}\right)^{-\frac{1}{2}}\right] e^{-\tau} \tag{59}$$

$$= \left(1 - \frac{2\tau}{N}\right)^{-\frac{N}{2}} \cdot e^{-\tau}, \tag{60}$$

where we have used the law of total expectation in (55); the conditional statistical independence of the elements of $\mathbf{u}(\mathsf{k}_b)$ (given k_b) in (58); the fact that $\{\mathsf{u}_n^2(\mathsf{k}_b) \sim \chi_1^2\}_{n=1}^N$, namely all the squared elements of $\mathbf{u}(\mathsf{k}_b)$ given k_b are chi-squared random variables with one degree of freedom; and, accordingly, that the moment generating function of a random variable $\mathsf{q} \sim \chi_1^2$ is $\mathbb{E}[e^{\widetilde{\tau}\mathsf{q}}] = (1 - 2\widetilde{\tau})^{-\frac{1}{2}}$, for some $\widetilde{\tau} < \frac{1}{2}$, in (59), where in our case $\widetilde{\tau} = \tau/N$, hence the condition on τ in (58) ∎

Equipped with Lemma 2, we now prove Lemma 1.

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

$$\widehat{\mathsf{k}}_b \triangleq \underset{m\in\mathcal{S}_{K_b}}{\arg\min} \left|\psi_N(\mathbf{y}, m)\right|. \tag{61}$$

In words, as $N \to \infty$, the error probability of (61) is governed by how far is $|\psi_N(\mathbf{y}, \mathsf{k}_b)|$ from zero, since from the TDC, $\nexists k \in \mathcal{S}_{K_b}/\mathsf{k}_b : \lim_{N\to\infty} |\psi_N(\mathbf{y}, k)| = 0$, whereas

$$\lim_{N\to\infty} \psi_N(\mathbf{y}, \mathsf{k}_b) = \mathbb{E}\left[\psi_N(\mathbf{y}, \mathsf{k}_b)\right] \tag{62}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\psi_N(\mathbf{y}, \mathsf{k}_b)|\mathsf{k}_b\right]\right] \tag{63}$$

$$= \frac{1}{N}\,\mathbb{E}\left[\mathbb{E}\left[\|\mathbf{u}(\mathsf{k}_b)\|_2^2|\mathsf{k}_b\right]\right] - 1 = 0, \tag{64}$$

where we have used (54), $\mathbf{u}(\mathsf{k}_b)|\mathsf{k}_b \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and (62) follows from the fact that $\mathbb{V}\mathrm{ar}(\psi_N(\mathbf{y}, \mathsf{k}_b)) = 2/N$, which can be shown in a similar fashion to (62)–(64).

Formally, the error probability of this estimator is given by,

$$\mathbb{P}\left[\widehat{\mathsf{k}}_b \neq \mathsf{k}_b\right] = \mathbb{P}\left[|\psi_N(\mathbf{y}, \mathsf{k}_b)| > \min_{m\in\mathcal{S}_{K_b}/\mathsf{k}_b} |\psi_N(\mathbf{y}, m)|\right]. \tag{65}$$

We now show that the probability that $\psi_N(\mathbf{y}, \mathsf{k}_b)$ is bounded away from zero decreases in the desired rate. Clearly, for any $a > 0$, we have

$$\mathbb{P}[|\psi_N(\mathbf{y}, \mathsf{k}_b)| > a] = \mathbb{P}[\psi_N(\mathbf{y}, \mathsf{k}_b) > a] \tag{66}$$

$$+ \mathbb{P}[\psi_N(\mathbf{y}, \mathsf{k}_b) < -a]\,. \tag{67}$$

Using the Chernoff bound, we have

$$\mathbb{P}[\psi_N(\mathbf{y}, \mathsf{k}_b) > a] \leq \mathbb{E}\left[e^{t\psi_N(\mathbf{y}, \mathsf{k}_b)}\right] e^{-ta} \triangleq \bar{B}_1(t, a), \tag{68}$$

$$\mathbb{P}[\psi_N(\mathbf{y}, \mathsf{k}_b) < -a] \leq \mathbb{E}\left[e^{-t\psi_N(\mathbf{y}, \mathsf{k}_b)}\right] e^{-ta} = \bar{B}_2(t, a). \tag{69}$$

Using Lemma 2, it follows that

$$\bar{B}_1(t, a) = \left(1 - \frac{2t}{N}\right)^{-\frac{N}{2}} \cdot e^{-t(1+a)}, \quad \forall t < \frac{N}{2}, \tag{70}$$

$$\bar{B}_2(t, a) = \left(1 + \frac{2t}{N}\right)^{-\frac{N}{2}} \cdot e^{-t(1-a)}, \quad \forall t > -\frac{N}{2}. \tag{71}$$

Minimizing $\bar{B}_1(t, a)$ and $\bar{B}_1(t, a)$ with respect to t and choosing $a = \log^{-1}(N)$, we obtain

$$\min_{t<\frac{N}{2}} \bar{B}_1(t, \log^{-1}(N)) = \left(\frac{1 + \log(N)}{\log(N)}\right)^{\frac{N}{2}} e^{-\frac{N}{2\log(N)}} \tag{72}$$

$$\triangleq \bar{B}_1^*(N), \tag{73}$$

$$\min_{t>-\frac{N}{2}} \bar{B}_2(t, \log^{-1}(N)) = \left(\frac{\log(N) - 1}{\log(N)}\right)^{\frac{N}{2}} e^{\frac{N}{2\log(N)}} \tag{74}$$

$$\triangleq \bar{B}_2^*(N). \tag{75}$$

Finally, since for any $\alpha \in \mathbb{R}_+$,

$$\lim_{N\to\infty} N^\alpha \bar{B}_1^*(N) = \lim_{N\to\infty} N^\alpha \bar{B}_2^*(N) = 0, \tag{76}$$

it follows that for any $\alpha \in \mathbb{R}_+$,

$$\mathbb{P}\left[|\psi_N(\mathbf{y}, \mathsf{k}_b)| > \frac{1}{\log(N)}\right] = \mathcal{O}\left(\frac{1}{N^\alpha}\right) \tag{77}$$

$$\implies \mathbb{P}\left[\widehat{\mathsf{k}}_b \neq \mathsf{k}_b\right] = \mathcal{O}\left(\frac{1}{N^\alpha}\right). \tag{78}$$

from which (13) follows, and this concludes the proof.