Joint Estimation of Carrier Frequency Offset and Channel Impulse Response for Linear Periodic Channels

Roee Shaked, *Student Member, IEEE*, Nir Shlezinger, *Student Member, IEEE*, and Ron Dabora, *Senior Member, IEEE*

***Abstract*— In many communications scenarios, the channel exhibits periodic characteristics, e.g., power line communications and interference-limited communications. Additionally, certain approximations for mobile radio channels over finite time intervals may result in periodic channel models. In this paper, we study pilot-aided joint estimation of the channel impulse response (CIR) and of the carrier frequency offset (CFO) for linear periodic channels, in which both the CIR and the noise statistics vary periodically in time. We first consider the joint maximum likelihood estimator (JMLE) for the CIR and the CFO, and discuss the practical drawbacks associated with this estimator. When the coefficients of the delay-Doppler spread function of the CIR are approximately sparse, we propose two estimation schemes with higher spectral efficiency and lower computational complexity compared with the JMLE, which are obtained by exploiting both the periodicity and the sparsity of the channel, without requiring a priori knowledge of the sparsity pattern. Finally, we study the design of pilot sequences aimed at improving the estimation performance in sparse periodic channels. Simulation studies corresponding to practical scenarios of the proposed estimators demonstrate that substantial benefits can be obtained by properly accounting for the sparsity and periodicity in the design of estimation schemes.**

***Index Terms*— Compressed channel sensing, cyclostationary signals, power line communications, JMLE, CFO estimation, CIR estimation, linear periodically time-varying.**

I. Introduction

MANY practical communications channels exhibit periodic characteristics: One important instance of a periodic communications channel is the narrowband (NB) power line communications (PLC) channel [1]. In such channels, periodicity is inherent in both the channel impulse response (CIR) and in the noise statistics, due to the periodicity of the mains voltage across the power grid [2]. Another major and frequently encountered communications scenario in which periodicity is inherent is interference-limited communications. In this communications scenario, periodicity follows since man-made signals are typically characterized by a cyclostationary statistical model [3, Sec. 5]. The class of interference-limited communications includes, for example, wireless communications, such as WiFi [4, Sec. 2] and 5G cellular [5], wireline communications [6], and cognitive radio [7]. Lastly, we note that the time-varying nature of channel impulse responses (CIRs) in mobile radio channels can sometimes be approximated using the complex exponential basis expansion model (CE-BEM) [8]–[11]. This modeling approach typically leads to a linear, periodically time-varying (LPTV) CIR. However, as wireless channels do not inherently have an LPTV CIR, then, using the CE-BEM leads to an *approximated CIR which is valid only for a time interval of a finite duration*, to which we refer as the *coherence interval (CI)* of the channel, and a new set of CIR parameters has to be estimated once in every CI. It thus follows that if the observation interval is shorter than the length of the CI, the resulting model can be approximated as a periodic communications channel for the duration of the CI.

Generally speaking, high data rate communications requires accurate knowledge of the CIR and of the carrier frequency offset (CFO), caused by instabilities of the oscillators [12, Sec. 1]. This need has motivated a large body of research works on the estimation of these parameters. Many CIR and CFO estimation schemes use an a-priori known pilot sequence (PS), incorporated into the transmitted signal, while the CIR and the CFO are modeled as *unknown deterministic parameters* [12, Sec. 3.1]. The work [13] designed a maximum likelihood (ML) scheme for the joint estimation of the CFO and of the CIR for linear time-invariant (LTI) channels with additive white Gaussian noise (AWGN). Joint ML estimation of the CIR and the CFO was studied in [12, Sec. 3.1.1], [14]-[16], for scenarios in which the signals are modulated using orthogonal frequency division multiplexing (OFDM); in [17] for multiple input-multiple output (MIMO) scenarios; and in [12, Sec. 4], [18] for MIMO-OFDM scenarios. The work [19] proposed a CFO estimation scheme for periodic channels *without estimating the CIR coefficients*. Yet, none of these works considered joint CFO and CIR estimation of linear periodically time-varying (LPTV) CIRs with additive cyclostationary Gaussian noise (ACGN).

Manuscript received January 22, 2017; revised May 29, 2017 and August 8, 2017; accepted August 8, 2017. Date of publication August 21, 2017; date of current version January 13, 2018. This work was supported by the Ministry of Economy of Israel through the Heron 5G Next Generation Consortium and by the Israel Science Foundation under Grant 1685/16. The associate editor coordinating the review of this paper and approving it for publication was L. Lampe. *(Corresponding author: Roee Shaked.)*

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Be'er-Sheva 8410501, Israel (e-mail: shroee@post.bgu.ac.il; nirshl@post.bgu.ac.il; ron@ee.bgu.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2017.2741943

0090-6778 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

PS design for joint CIR and CFO estimation over scalar LTI channels with AWGN was studied in [20], where the objective was to minimize the Cramér-Rao bound (CRB) for the mean-square error (MSE) in estimating the CIR. PS design for joint CIR and CFO estimation was also studied in [17], which considered LTI multiple input-multiple output (MIMO) channels with AWGN. However, neither [17] nor [20] considered time variations of the CIR or ACGN. The work [11] studied PS design and pilot placement for orthogonal frequency division multiplexing (OFDM) communications over channels for which the time variations of the CIR were modeled using a CE-BEM, yet, [11] did not consider CFO.

A major issue in the estimation of CIRs is the need to estimate a large number of parameters. As evident from [17] and [21], the CRB for estimating the coefficients of the CIR has a linear dependence on the number of CIR coefficients estimated. Consequently, reducing the number of estimated CIR coefficients improves estimation performance. A common approach to reduce the number of estimated parameters is to assume that the CIR is sparse, and employ compressed sensing (CS) techniques [22], [23]. Indeed, in [21] and [24]–[26], it was shown that applying CS-based methods results in an improved performance for CIR estimation. An additional important benefit of CS-based methods is that they require a shorter PS compared to conventional least squares (LS)-based estimators [24]. Consequently, applying CS-based methods can improve the spectral efficiency, as a larger duration of time can be allocated for data transmission instead of pilots transmission. This improvement in spectral efficiency is especially relevant to communications scenarios in which the coefficients of the model representing the CIR can be considered constants only for a limited duration, as is the case when applying the complex exponential basis expansion model (CE-BEM) to wireless channels. The works [27], [28] proposed a CS-based joint CIR and CFO estimator for slow-fading LTI channels with AWGN. In particular, [27] proposed an iterative algorithm which updates the CFO ML estimate at each iteration using the expectation-maximization algorithm, followed by an update of the CIR estimate using the subspace pursuit CS algorithm. However, the estimator proposed by [27] assumes that the number of non-zero CIR coefficients is *a-priori known*. The work [28] proposed to estimate the CFO value using block orthogonal matching pursuit after quantizing the CFO, followed by LS estimation of the CIR. Yet, [28] assumed that the number of observations *is larger* than the number of estimated parameters. In the works [8]–[10] a time-varying CIR is modeled using a small number of complex exponentials, which results in the channel having a *sparse CE-BEM*. In particular, we note that [8, Sec. III.B] and [9, Sec. 4.1] proposed an algorithm for detecting the frequencies of the CE-BEM based on cyclic moments *without utilizing* CS techniques. Lastly, the work [29] designed an estimator for the CIR coefficients in LPTV channels with temporally uncorrelated ACGN, whose variance is known at the receiver for each time instance. The estimator in [29] also assumed that the CFO and the sparsity pattern of the discrete Fourier series (DFS) of the CIR coefficients are *a-priori known*. To the best of our knowledge, a joint estimator of LPTV CIR coefficients and CFO in the presence of correlated ACGN, with or without the use of CS techniques, has not yet been derived.

A. Main Contributions

In this paper we study the joint estimation of the CIR coefficients and of the CFO for LPTV channels with ACGN, where the CIR coefficients and CFO are assumed deterministic and unknown. The main contributions of this work are as follows:

- We first consider the joint ML estimator (JMLE) for both the CIR and the CFO for transmission over LPTV channels with ACGN. We note that the number of coefficients of an LPTV CIR is equal to the product of the length of the memory times the length of the period, and is typically much larger than the number of coefficients of an LTI CIR with the same memory length. Coupled with the memory of the noise, we assert that the JMLE for this problem has a low spectral efficiency and a very high computational complexity.
- The above observation motivated us to apply sparsity–based techniques, aimed at improving the complexity vs. performance tradeoff. As a first step, we analyzed the relationship between the CFO and the sparsity pattern of the DFS coefficients of the CIR. Our analysis shows, for the first time, that *when the CFO is unknown, then the sparsity pattern is also unknown*. We thus conclude that for an unknown CFO it is not possible to apply estimation schemes based on a-priori knowledge of the sparsity pattern.
- This conclusion led us to propose two CS-based joint ML estimator (JMLE)-oriented schemes which utilize the sparse nature of the coefficients of the DFS of the CIR, without requiring a-priori knowledge of the sparsity pattern itself. The proposed schemes require less pilot symbols compared to the JMLE, thereby increasing the spectral efficiency for data transmission, compared to that of the JMLE. Furthermore, the proposed schemes require the estimation of a subset of channel coefficients compared to the JMLE, and consequently, their computational complexity is smaller than that of the JMLE. In fact, our numerical study of practical scenarios demonstrates that the proposed schemes can outperform both the JMLE as well as an adaptation of the algorithm proposed in [13] to periodic channels, in terms of estimation error, spectral efficiency, and computational complexity, simultaneously.
- Lastly, we study the design of PSs for improving the performance of CFO and CIR estimation over periodic channels with sparse DFS coefficients, and propose a new approach for designing PSs. Simulation results show that PSs designed using the new approach achieve better performance in practical scenarios compared to commonly used PSs.

The rest of this paper is organized as follows: Section II states the problem formulation and details the pilot-aided JMLE for estimating the CFO and the CIR coefficients; Section III introduces the two JMLE-oriented estimators which

use sparsity based techniques; Section IV considers the design of the PS; Section V presents a numerical study of the proposed estimators; Lastly, Section VI provides some concluding remarks.

II. Preliminaries

A. Mathematical Notations and Definitions

We denote the sets of non-negative integers, integers, and complex numbers by $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{C}$, respectively. $|\mathcal{X}|$ denotes the cardinality of the set $\mathcal{X}$. Column vectors are denoted with lower-case boldface letters, e.g., $\mathbf{x}$; the k-th element ($k \geq 0$) of a vector $\mathbf{x}$ is denoted with $[\mathbf{x}]_k$. Matrices are denoted with upper-case boldface letters, e.g., $\mathbf{X}$, where the element at the k-th row and the l-th column of $\mathbf{X}$ is denoted by $[\mathbf{X}]_{k,l}$. $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, $\mathbf{0}_{N\times M}$ denotes an $N \times M$ matrix of zeros, $\delta[\cdot]$ denotes the Kronecker delta function, and $\otimes$ denotes the Kronecker product. Hermitian transpose, transpose, complex conjugate, Euclidean norm, and stochastic expectation are denoted by $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^*$, $\|\cdot\|$, and $\mathbb{E}\{\cdot\}$, respectively. $\rho\{\mathbf{X}\}$ denotes the largest singular value of the matrix $\mathbf{X}$, or equivalently, the square root of the largest eigenvalue of $\mathbf{X}^H\mathbf{X}$. For $\mathbf{X}$ a projection matrix onto the space $\mathcal{X} \subset \mathbb{C}^N$, we denote the projection matrix onto the subspace $\mathcal{X}^\perp \subset \mathbb{C}^N$, orthogonal to $\mathcal{X}$ with $\mathbf{X}^\perp \triangleq \mathbf{I}_N - \mathbf{X}$. $\mathrm{Re}\{x\}$ and $\mathrm{Im}\{x\}$ denote the real part and the imaginary part of $x \in \mathbb{C}$, respectively, $\langle x \rangle$ denotes the fractional part of a real number x s.t. $\langle x \rangle \in (-1/2, 1/2] \triangleq I_{1/2}$, $\lfloor x \rfloor$ denotes the largest integer not greater than x, and for $x \in \mathbb{Z}$, $((x))_m \in \{0, 1, ..., m-1\}$ denotes the remainder of x when divided by $m \in \mathbb{N}$. $\mathcal{CN}$ denotes the proper-complex (PC) Gaussian distribution, and $D\left(f_1||f_2\right)$ denotes the Kullback-Leibler divergence (KLD) between two probability measures f_1 and f_2, defined in [30, eq. (17.18)]. Lastly, recall the following definition:

Definition 1: A discrete-time PC process $x[n]$ is said to be wide-sense cyclostationary (WSCS) if both its mean value and autocorrelation function are periodic with a period of $N_0 \in \mathbb{N}$, i.e., $\mathbb{E}\{x[n]\} = \mathbb{E}\{x[n+N_0]\}$, and $c_x[n,l] \triangleq \mathbb{E}\{x[n]x^*[n+l]\} = c_x[n+N_0,l]$, see [3, Sec. 3.5].

B. Signal Model and Problem Formulation

We consider a discrete-time scalar, finite-memory, periodic channel, modeled as an LPTV filter with ACGN. Let $L_{\text{noise}}, N_{\text{noise}} \in \mathbb{N}$ denote *the length and the period of the temporal correlation of the noise*, respectively. Let $w[n]$ denote the PC zero-mean ACGN with autocorrelation function $c_w[n,l] \triangleq \mathbb{E}\{w[n]w^*[n+l]\}$. Thus, $c_w[n,l] = c_w[n+N_{\text{noise}},l]$, $\forall n,l \in \mathbb{Z}$, and $c_w[n,l] = 0$ for all $|l| \geq L_{\text{noise}}$. Next, let $L_{\text{ch}}, N_{\text{ch}} \in \mathbb{N}$ denote *the length and the period of the CIR*, respectively. Let $g[n,l]$ denote the CIR of the causal LPTV channel, namely, $\forall n, l_0 \in \mathbb{Z}$, $g[n, n-l_0]$ is the output of the channel for input $\delta[n-l_0]$, when the noise is zero. The CIR $g[n,l]$ is assumed to be deterministic and unknown to the receiver. The periodicity of the CIR implies that $g[n,l] = g[n+N_{\text{ch}},l]$, $\forall n,l \in \mathbb{Z}$, while the causality and the finite-memory imply that $g[n,l] \neq 0$ only for $0 \leq l < L_{\text{ch}}$. Set N_0 to be the least common multiple of N_{ch} and N_{noise}, s.t. $N_0 > \max\{L_{\text{noise}}, L_{\text{ch}}\}$. We hereafter refer to N_0 as the *period of the channel*. N_0 is assumed a-priori known to the receiver.[1] Let $\nu \in I_{1/2}$ denote the CFO between the carrier frequencies at the transmitter and at the receiver, normalized to the sampling rate, and let ϕ_0 denote the corresponding phase offset at time $n = 0$, both measured at the channel output. Lastly, we use $x[n]$ and $y[n]$ to denote the channel input and the channel output at time $n \in \mathbb{Z}$, respectively. Assuming no sampling frequency offset between the transmitter and the receiver, the input-output relationship is given by

$$y[n] = e^{j(2\pi\nu n+\phi_0)} \sum_{l=0}^{L_{\text{ch}}-1} g[n,l]x[n-l] + w[n], \quad n \in \mathbb{Z}. \tag{1}$$

Define the parameter $\alpha \triangleq \langle N_0 \cdot \nu \rangle$, which represents the phase offset accumulated over a single channel period, and define $h[n,l] \triangleq e^{j\left(2\pi\nu((n))_{N_0}+\phi_0\right)} g[n,l]$. The filter $h[n,l]$ represents the phase shifted CIR (recall that the phase shift is unknown). Observe that $h[n,l]$ inherits the periodicity and finite-memory of $g[n,l]$, i.e., $h[n,l] = h[n+N_0,l]$ and $h[n,l] \neq 0$ only for $0 \leq l < L_{\text{ch}}$. Substituting the definition of $h[n,l]$ into (1) yields

$$y[n] = e^{j2\pi\alpha\left\lfloor\frac{n}{N_0}\right\rfloor} \sum_{l=0}^{L_{\text{ch}}-1} h[n,l]x[n-l] + w[n]. \tag{2}$$

We note that a similar input-output relationship was assumed in [19], which did not consider CIR estimation. Since the CIR is periodic, it has a DFS representation [31, Ch. 8.1] with DFS coefficients $\underset{\sim}{h}[k,l]$ given by:

$$\underset{\sim}{h}[k,l] \triangleq \frac{1}{\sqrt{N_0}} \sum_{n=0}^{N_0-1} h[n,l]e^{-j\frac{2\pi}{N_0}nk}, \tag{3}$$

for $k \in \{0, 1, ..., N_0-1\} \triangleq \mathcal{N}_0$, $l \in \{0, 1, ..., L_{\text{ch}}-1\} \triangleq \mathcal{L}_{\text{ch}}$. Hereafter, we refer to $\underset{\sim}{h}[k,l]$ as the delay-Doppler spreading function (DDSF), as in [21] and [26]. Note that $h[n,l]$ can be obtained from $\underset{\sim}{h}[k,l]$ via the inverse DFS (IDFS): $h[n,l] = \frac{1}{\sqrt{N_0}}\sum_{k=0}^{N_0-1} \underset{\sim}{h}[k,l]e^{j\frac{2\pi}{N_0}kn}$, $n \in \mathbb{Z}$, $l \in \mathcal{L}_{\text{ch}}$. Substituting the IDFS of $\underset{\sim}{h}[k,l]$ into (2) yields the input-output relationship

$$y[n] = \sum_{l=0}^{L_{\text{ch}}-1} \frac{x[n-l]}{\sqrt{N_0}} \sum_{k=0}^{N_0-1} e^{j\frac{2\pi}{N_0}nk} e^{j2\pi\alpha\left\lfloor\frac{n}{N_0}\right\rfloor} \underset{\sim}{h}[k,l] + w[n]. \tag{4}$$

Note that the term $e^{j2\pi\alpha\left\lfloor\frac{n}{N_0}\right\rfloor}$ in (4) is a non-periodic term in an otherwise periodic CIR.

Our objective is to design a scheme for estimating the deterministic unknown parameters $\left(\alpha, \{\underset{\sim}{h}[k,l]\}_{k\in\mathcal{N}_0, l\in\mathcal{L}_{\text{ch}}}\right)$ from an observed sequence of channel outputs. It is assumed the receiver knows the statistical characterization of the noise.

[1] Assuming that N_0 is a-priori known is a reasonable assumption in many periodic communications scenarios, e.g., narrowband (NB)-power line communications (PLC) [2] and interference-limited communications [4], and such knowledge was also assumed in previous works which studied estimation of time-varying channels modeled using periodic functions, e.g., [11], [21]. We note that inaccurate knowledge of the period will necessarily degrade the performance of any estimation scheme operating in a periodic scenario.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE JMLE (7)

	Term	Complex $\times$	Complex $+$
(8*a*)	$\mathbf{C}_{\mathbf{w}}^{-1}$	$\mathcal{O}\left(N_{\text{obs}}^3\right)$	$\mathcal{O}\left(N_{\text{obs}}^3\right)$
(8*b*)	$\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}$	N_{obs}^2	$N_{\text{obs}}(N_{\text{obs}}-1)$
(8*c*)	$\mathbf{SF}$	$N_{\text{p}}N_0N_{\text{CIR}}$	
(8*d*)	$(\mathbf{SF})^H(\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y})$	$N_{\text{p}}N_0N_{\text{CIR}}$	$N_{\text{p}}(N_0-1)N_{\text{CIR}}$
(8*e*)	$(\mathbf{SF})^H\mathbf{C}_{\mathbf{w}}^{-1}(\mathbf{SF})$	$N_{\text{p}}N_{\text{CIR}}(N_{\text{obs}}+\frac{N_{\text{p}}N_{\text{CIR}}+1}{2})N_0$	$N_{\text{p}}N_{\text{CIR}}(N_{\text{obs}}+\frac{N_{\text{p}}N_{\text{CIR}}+1}{2})(N_0-1)$
(8*f*)	$\mathbf{v}_{(\alpha)}\triangleq\mathbf{\Phi}_{(\alpha)}^H\left((\mathbf{SF})^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}\right)$	$N_{\text{CIR}}\cdot\mathcal{O}(T\log(N_{\text{p}}))$	$N_{\text{CIR}}\cdot\mathcal{O}(T\log(N_{\text{p}}))$
(8*g*)	$\mathbf{\Gamma}_{(\alpha)}\triangleq\left(\mathbf{\Phi}_{(\alpha)}^H\left((\mathbf{SF})^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{SF}\right)\mathbf{\Phi}_{(\alpha)}\right)^{-1}$	$T\left(\mathcal{O}\left(N_{\text{CIR}}^3\right)+\frac{N_{\text{CIR}}(N_{\text{CIR}}+1)N_{\text{p}}(N_{\text{p}}-1)}{2}\right)$	$T\left(\mathcal{O}\left(N_{\text{CIR}}^3\right)+\frac{N_{\text{CIR}}(N_{\text{CIR}}+1)N_{\text{p}}^2-1}{2}\right)$
(8*h*)	$\mathbf{v}_{(\alpha)}^H\mathbf{\Gamma}_{(\alpha)}\mathbf{v}_{(\alpha)}$	$2T\cdot N_{\text{CIR}}^2$	$T\left(N_{\text{CIR}}^2-1\right)$
(8*i*)	$\mathbf{\Gamma}_{\left(\hat{\alpha}^{\text{JML}}\right)}\mathbf{v}_{\left(\hat{\alpha}^{\text{JML}}\right)}$	N_{CIR}^2	$N_{\text{CIR}}(N_{\text{CIR}}-1)$

C. The JMLE

In this section, we derive the JMLE for the CFO and the CIR. Let $N_{\text{CIR}}\triangleq N_0L_{\text{ch}}$ denote the number of CIR coefficients and let N_{obs} denote the number of observations used for computing the estimate. The channel input consists of a predefined PS of length $N_{\text{seq}}=N_{\text{obs}}+L_{\text{ch}}-1$, $\{s[n]\}_{n=-L_{\text{ch}}+1}^{N_{\text{obs}}-1}$. For simplicity, we assume that N_{obs} is an integer multiple of the channel period, $N_{\text{obs}}=N_{\text{p}}N_0$, $N_{\text{p}}\in\mathbb{N}$. When the PS $s[n]$ is transmitted, the channel outputs at time indexes $n\in\{0,1,..,N_{\text{obs}}-1\}\triangleq\mathcal{N}_{\text{obs}}$ are obtained from (4) with $x[n]=s[n]$. Next, define the vectors $\mathbf{y},\mathbf{w}\in\mathbb{C}^{N_{\text{obs}}}$ s.t. $\left[\mathbf{y}\right]_n=y[n]$, $[\mathbf{w}]_n=w[n]$, $n\in\mathcal{N}_{\text{obs}}$, and define the $N_{\text{obs}}\times N_{\text{obs}}$ matrix $\mathbf{C}_{\mathbf{w}}\triangleq\mathbb{E}\left\{\mathbf{w}\mathbf{w}^H\right\}$ s.t. $[\mathbf{C}_{\mathbf{w}}]_{n_1,n_2}=c_w[n_1,n_2-n_1]$. Additionally, define the DDSF coefficients vector $\underset{\sim}{\mathbf{h}}\in\mathbb{C}^{N_{\text{CIR}}}$ s.t. $[\underset{\sim}{\mathbf{h}}]_{kL_{\text{ch}}+l}=\underset{\sim}{h}[k,l]$, $k\in\mathcal{N}_0$, $l\in\mathcal{L}_{\text{ch}}$, and the PS matrix $\mathbf{S}\in\mathbb{C}^{N_{\text{obs}}\times N_{\text{p}}N_{\text{CIR}}}$ s.t. $[\mathbf{S}]_{n,l}=s\left[n-((l))_{L_{\text{ch}}}\right]$, for $n\in\mathcal{N}_{\text{obs}}$, $nL_{\text{ch}}\le l<(n+1)L_{\text{ch}}$, and $[\mathbf{S}]_{n,l}=0$ otherwise. Let $\mathbf{F}_{N_0}^{-1}$ be the $N_0\times N_0$ IDFS matrix, i.e., $\left[\mathbf{F}_{N_0}^{-1}\right]_{n,k}=\frac{1}{\sqrt{N_0}}e^{j\frac{2\pi}{N_0}kn}$, and define the $N_{\text{p}}N_{\text{CIR}}\times N_{\text{p}}N_{\text{CIR}}$ matrix $\mathbf{F}\triangleq\mathbf{I}_{N_{\text{p}}}\otimes\mathbf{F}_{N_0}^{-1}\otimes\mathbf{I}_{L_{\text{ch}}}$. Lastly, define the matrix $\mathbf{\Phi}_{(\alpha)}\in\mathbb{C}^{N_{\text{p}}N_{\text{CIR}}\times N_{\text{CIR}}}$ s.t. $\mathbf{\Phi}_{(\alpha)}\triangleq[1,e^{j2\pi\alpha},...,e^{j2\pi(N_{\text{p}}-1)\alpha}]^T\otimes\mathbf{I}_{N_{\text{CIR}}}$, the matrix $\mathbf{H}_{(\alpha)}\triangleq\mathbf{SF}\mathbf{\Phi}_{(\alpha)}$, and the matrix $\mathbf{G}_{(\alpha)}\triangleq\left(\mathbf{H}_{(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}_{(\alpha)}\right)^{-1}$. Applying the above definitions to the relationship (4), we arrive at the following multivariate input-output relationship:

$$\mathbf{y}=\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}+\mathbf{w},\qquad\mathbf{y}\sim\mathcal{CN}\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}},\mathbf{C}_{\mathbf{w}}\right).\tag{5}$$

Note that the observations model in (5) is linear in $\underset{\sim}{\mathbf{h}}$ and non-linear in α, and since $c_w[n,l]$ is assumed to be known, the receiver can construct $\mathbf{C}_{\mathbf{w}}$. The JMLE of α and $\underset{\sim}{\mathbf{h}}$ based on $\mathbf{y}$, denoted $(\hat{\alpha}^{\text{JML}},\hat{\underset{\sim}{\mathbf{h}}})$, is given by the solution to the following weighted LS (WLS) problem [32, Ch. 8.9]:

$$\left(\hat{\alpha}^{\text{JML}},\hat{\underset{\sim}{\mathbf{h}}}^{\text{JML}}\right)=\underset{\alpha\in I_{1/2},\underset{\sim}{\mathbf{h}}\in\mathbb{C}^{N_{\text{CIR}}}}{\text{argmin}}\left\{(\mathbf{y}-\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}})^H\mathbf{C}_{\mathbf{w}}^{-1}(\mathbf{y}-\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}})\right\}.\tag{6}$$

When the number of observations is *larger* than the number of estimated parameters, i.e., $N_{\text{p}}>L_{\text{ch}}$, and the matrix $\mathbf{H}_{(\alpha)}$ is of full rank, then the JMLE can be obtained from (6) following the steps as in [32, Ch. 8.9]: First, the CFO is estimated via

$$\hat{\alpha}^{\text{JML}}=\underset{\alpha\in I_{1/2}}{\text{argmax}}\left\{\mathbf{y}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}_{(\alpha)}\mathbf{G}_{(\alpha)}\mathbf{H}_{(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}\right\},\tag{7a}$$

and then, the CIR is estimated via

$$\hat{\underset{\sim}{\mathbf{h}}}_{(\hat{\alpha}^{\text{JML}})}^{\text{JML}}=\mathbf{G}_{(\hat{\alpha}^{\text{JML}})}\mathbf{H}_{(\hat{\alpha}^{\text{JML}})}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}.\tag{7b}$$

Note that when the unknown channel $g[n,l]$ is LTI and the noise is AWGN, i.e., $N_0=1$ and $\mathbf{C}_{\mathbf{w}}=\sigma^2\mathbf{I}_{N_{\text{obs}}}$, then, from the definitions of $\mathbf{F}$ and α, we obtain $\mathbf{F}=\mathbf{I}_{N_{\text{obs}}L_{\text{ch}}}$ and $\alpha=\nu$. In this case, the JMLE in (7) coincides with the JMLE proposed in [13].

Complexity Analysis: The complexity analysis of the JMLE, measured via the number of complex multiplications and the number of complex additions, is summarized in Table I. In the following we briefly overview the considerations used for obtaining the computational complexity expressions appearing in Table I: In the analysis it is assumed that α is estimated by evaluating the expression in Eqn. (7a) at T test points, where T is selected to satisfy the required estimation accuracy. Typically, this results in a value of T which is large compared to N_{CIR} and N_{obs}. We note that the complexities of computing (8c), (8d), and (8f) are obtained by using the sparse structure of the matrices $\mathbf{S}$, $\mathbf{F}$, and $\mathbf{\Phi}_{(\alpha)}$; in determining the complexity of the expressions in (8e) and (8g), we used the fact that, as these matrices are Hermitian, it is sufficient to compute only their upper triangular entries. The complexities of computing the expressions (8f)-(8h) include a multiplicative factor of T as these computations are intermediate steps in the evaluation of (7a), which is computed for T test values. Finally, (8i) corresponds to evaluating (7b), which is computed only once. From the complexity analysis we conclude that the complexity of calculating the estimate (7) is dominated by T inversions of an $N_{\text{CIR}}\times N_{\text{CIR}}$ complex matrix stated in (8g), and thus, the complexity for computing the estimate (7) is on the order of $\mathcal{O}\left(T\cdot N_{\text{CIR}}^3\right)$.

In summary, the JMLE has two major drawbacks when considering its application to practical scenarios:

1) **Low spectral efficiency**: As a total of $N_{\mathrm{CIR}}+1$ parameters, consisting of $N_{\mathrm{CIR}}=L_{\mathrm{ch}}\cdot N_0$ CIR coefficients and the CFO, are estimated based on $N_{\mathrm{obs}}=N_{\mathrm{p}}\cdot N_0$ observations via (7), then N_{p} must be larger than L_{ch}. Moreover, in order to achieve a high accuracy, N_{p} will have to be much larger than L_{ch}. A large N_{p} implies that a large number of pilot symbols is needed, which results in a low spectral efficiency.
2) **High computational complexity**: The computation of the estimate has a very high computational complexity, which is dominated by the complexity of the grid search in (7a).

We next propose two modifications of the JMLE which improve the spectral efficiency and reduce the computational complexity, based on sparsity of the DDSF vector $\underset{\sim}{\mathbf{h}}$.

III. Estimation Schemes Based on Utilizing Sparsity

Intuitively, the sparsity assumption implies that "most of the information" about the CIR is contained in a small number of its DDSF coefficients, and therefore $\underset{\sim}{\mathbf{h}}$ can be *approximated as a sparse vector*. Specifically, we assume that there exists a set of indexes $\mathcal{K}\subseteq\{0,1,\ldots,N_{\mathrm{CIR}}-1\}\triangleq\mathcal{N}_{\mathrm{CIR}}$, s.t. $[\underset{\sim}{\mathbf{h}}]_k\approx 0$ for each $k\notin\mathcal{K}$. The set $\mathcal{K}$ is referred to in the following as the *sparsity pattern*. A similar type of sparsity assumption was used in the estimation of sparse LTI CIRs in [24] and [25], the estimation of sparse time-varying CIRs in [21] and [26], and also for the estimation of sparse LPTV CIRs in [29]. It is emphasized that differently from the current work, [29] assumed that *the set $\mathcal{K}$ is a-priori known*, and restricted the structure of the sparsity to satisfy that at some values of $k\in\mathcal{N}_0$, $\underset{\sim}{h}[k,l]=0$ for all lags $l\in\mathcal{L}_{\mathrm{ch}}$. In practice, wireless communications channels typically exhibit sparsity w.r.t. specific lags corresponding to the delays of the various propagation paths [25, Sec. II], i.e., unlike [29], $\underset{\sim}{h}[k,l]$ will typically be approximately zero for specific values of $l\in\mathcal{L}_{\mathrm{ch}}$, for all $k\in\mathcal{N}_0$, and in general, $\underset{\sim}{h}[k,l]$ may be equal to zero for any index pair (k,l), $k\in\mathcal{N}_0$, $l\in\mathcal{L}_{\mathrm{ch}}$ [21, Sec. II.C]. Furthermore, as we show in the following, when the CFO is unknown, the sparsity pattern $\mathcal{K}$ *cannot be determined a-priori*. To prove this statement, let $\underset{\sim}{g}[k,l]$ denote the DFS of $g[n,l]$ for a given lag l, i.e., $\underset{\sim}{g}[k,l]$ is the DDSF of the CIR in the absence of CFO and phase offset. Recall that $\underset{\sim}{\mathbf{h}}$ is defined as the stacking of $\underset{\sim}{h}[k,l]$, which is the DDSF of the LPTV filter $h[n,l]$ computed via (3). The relationship between the non-zero elements of $\underset{\sim}{h}[k,l]$ and of $\underset{\sim}{g}[k,l]$ is stated in the following lemma:

Lemma 1: The coefficient $\underset{\sim}{h}[k,l]$ is equal to zero if and only if the CFO ν satisfies one of the following two conditions:

- *Condition 1: $N_0\nu\in\mathbb{Z}$ (equivalently, $\alpha=0$), and $\underset{\sim}{g}\left[((k-N_0\nu))_{N_0},l\right]=0$.*
- *Condition 2: $N_0\nu\notin\mathbb{Z}$, and*

$$\sum_{k'=0}^{N_0-1}\underset{\sim}{g}\left[k',l\right]\frac{\sin\left(\pi\left(k-k'-N_0\nu\right)\right)}{\sin\left(\frac{\pi\left(k-k'-N_0\nu\right)}{N_0}\right)}\cdot e^{-j\pi\frac{(N_0-1)}{N_0}(k-k'-N_0\nu)}=0.$$

[A proof is given in Appendix A]

Lemma 1 implies that the *sparsity pattern $\mathcal{K}$ depends on the unknown parameter ν,* hence, when ν is unknown, *the set $\mathcal{K}$ cannot be a-priori known,* and it has to be estimated. Therefore, we propose an estimation scheme which accounts for the sparsity of $\underset{\sim}{\mathbf{h}}$ by applying sparsity pattern estimation. The proposed estimator achieves improved performance in the joint estimation of the CFO and CIR, coupled with *increased spectral efficiency and low computational complexity*.

A. Approximated JMLE With Sparsity Pattern Estimation

It follows from (5) that the observations vector $\mathbf{y}$ is given by the product of the approximately sparse vector $\underset{\sim}{\mathbf{h}}$ and the matrix $\mathbf{H}_{(\alpha)}$, *which is a function of the unknown parameter α*, corrupted by the additive noise vector $\mathbf{w}$. The dependence of $\mathbf{H}_{(\alpha)}$ on α does not allow the direct implementation of existing sparsity estimation algorithms, e.g., [22, Ch. 1.6], [23], which require to model the observations as a product between a *known* matrix and an unknown vector, possibly corrupted by an additive noise vector. To circumvent this issue, consider the explicit relationship (4), and note that the DDSF coefficients $\underset{\sim}{h}[k,l]$ are multiplied by a complex exponential $e^{j2\pi\alpha\left\lfloor\frac{n}{N_0}\right\rfloor}$, whose phase increases by $2\pi\alpha$ with each sequence of N_0 observations. With this in mind, we define $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)\in\mathbb{C}^{N_{\mathrm{p}}N_{\mathrm{CIR}}}$ s.t. $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)\triangleq\mathbf{\Phi}_{(\alpha)}\underset{\sim}{\mathbf{h}}$, and the set $\mathcal{N}_{\mathrm{p}}\triangleq\{0,\ldots,N_{\mathrm{p}}-1\}$. From the definition of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ it follows that if at some index $k\in\mathcal{N}_{\mathrm{CIR}}$, $[\underset{\sim}{\mathbf{h}}]_k$ has a large magnitude, then $[\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)]_{k+pN_{\mathrm{CIR}}}$ also has a large magnitude, $\forall p\in\mathcal{N}_{\mathrm{p}}$. The relationship between the sparsity patterns of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ and of $\underset{\sim}{\mathbf{h}}$ is characterized in the following lemma:

Lemma 2: Let $\mathcal{K}$ denote the sparsity pattern of $\underset{\sim}{\mathbf{h}}$. The sparsity pattern of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ is given by

$$\underset{\sim}{\mathcal{K}}=\left\{\underset{\sim}{k}\in\mathbb{N}|\underset{\sim}{k}=pN_{\mathrm{CIR}}+k,\,p\in\mathcal{N}_{\mathrm{p}},k\in\mathcal{K}\right\}. \tag{9}$$

Proof: By definition $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)=\mathbf{\Phi}_{(\alpha)}\underset{\sim}{\mathbf{h}}$, hence, $[\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)]_{pN_{\mathrm{CIR}}+k}\triangleq e^{j2\pi\alpha p}[\underset{\sim}{\mathbf{h}}]_k$, $\forall p\in\mathcal{N}_{\mathrm{p}}$, $k\in\mathcal{N}_{\mathrm{CIR}}$, $\alpha\in I_{1/2}$, and thus, $\left|[\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)]_{pN_{\mathrm{CIR}}+k}\right|=\left|e^{j2\pi p\alpha}[\underset{\sim}{\mathbf{h}}]_k\right|=\left|[\underset{\sim}{\mathbf{h}}]_k\right|$. It thus follows that $[\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)]_{\underset{\sim}{k}}=0$ if and only if $[\underset{\sim}{\mathbf{h}}]_{((\underset{\sim}{k}))_{N_{\mathrm{CIR}}}}=0$, i.e., $\underset{\sim}{k}\in\underset{\sim}{\mathcal{K}}$, if and only if $\left(\left(\underset{\sim}{k}\right)\right)_{N_{\mathrm{CIR}}}\in\mathcal{K}$. □

Consequently, the sparsity pattern of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ consists of the sparsity pattern of $\underset{\sim}{\mathbf{h}}$ replicated N_{p} times, once in every subsequent N_{CIR} elements, thus $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ exhibits *group sparsity* [33], [34], and $\mathcal{K}=\underset{\sim}{\mathcal{K}}\cap\mathcal{N}_{\mathrm{CIR}}$. Using the definition of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$, Eqn. (5) can be written as

$$\mathbf{y}=\mathbf{SF}\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)+\mathbf{w}. \tag{10}$$

In the terminology used in CS problems, the observations $\mathbf{y}$ in (10) correspond to the product of a *known* sensing matrix $\mathbf{SF}$ with an unknown group-sparse vector $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$, corrupted by the additive noise $\mathbf{w}$. Based on the model (10), we propose a group sparsity estimation scheme using the adaptation of the basis pursuit denoising (BPDN) algorithm [23] to signals possessing structured sparsity patterns, as detailed in [33] and [34]. The sparsity estimation scheme consists of three steps:

1) Fix a positive σ^2 such that $\sigma^2\geq\mathbb{E}\left\{\|\mathbf{w}\|^2\right\}$.
2) Using the BPDN variation detailed in [33] and [34], the recovery of the group sparse vector $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ in (10) is

stated as the solution to following convex minimization problem:

$$\underset{\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)\in\mathbb{C}^{N_{\mathrm{p}}N_{\mathrm{CIR}}}}{\operatorname{argmin}} \sum_{k=0}^{N_{\mathrm{CIR}}-1}\sqrt{\sum_{p=0}^{N_{\mathrm{p}}-1}\left|[\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)]_{pN_{\mathrm{CIR}}+k}\right|^2} \quad \text{s.t.} \quad ||\mathbf{y}-\mathbf{S}\mathbf{F}\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)||^2\le\sigma^2. \tag{11}$$

3) Let $\underset{\sim}{\mathcal{K}}$ be the support of $\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)$. The sparsity pattern estimate is $\mathcal{K}=\underset{\sim}{\mathcal{K}}\cap\mathcal{N}_{\mathrm{CIR}}$.

Note that the BPDN estimate $\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)$ is used only for identifying the sparsity pattern of $\underset{\sim}{\mathbf{h}}$, and not for estimating the values of the unknown parameters α and $\underset{\sim}{\mathbf{h}}$, as $\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)$ is likely to be an inaccurate estimate of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$. This is due to two main factors: The first is that $\underset{\sim}{\mathbf{h}}$ is only approximately sparse, and the second is that the BPDN algorithm is ignorant of the underlying structure of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ and its relationship with the structure of $\underset{\sim}{\mathbf{h}}$ and with α, stated in Lemma 2. Consequently, the BPDN estimates a larger number of parameters, which in turn degrades the estimation accuracy.

Unlike the work [29], the proposed algorithm does not impose constraints on the sparsity pattern, and any index $kL_{\mathrm{ch}}+l$, $k\in\mathcal{N}_0$, $l\in\mathcal{L}_{\mathrm{ch}}$ may belong to the set $\mathcal{K}$. Accordingly, the algorithm is capable of utilizing sparsity both in the time-lag domain (i.e., in $\mathcal{L}_{\mathrm{ch}}$) and in the Doppler domain (i.e., in $\mathcal{N}_0$). Furthermore, the proposed algorithm may serve as an alternative to the algorithm of [8, Sec. III.B.] and [9, Sec. 4.1] for detecting the frequencies of the CE-BEM.

Having estimated the sparsity pattern $\mathcal{K}$, we next state the approximation of the JMLE under the sparsity assumption. Let $\underset{\sim}{\mathbf{h}}_{\mathrm{r}}\in\mathbb{C}^{|\mathcal{K}|}$ denote the vector of DDSF coefficients of *reduced size*, obtained by taking only the entries of $\underset{\sim}{\mathbf{h}}$ whose indexes are in the estimated sparsity pattern $\mathcal{K}$. Accordingly, we omit from $\mathbf{\Phi}_{(\alpha)}$ the columns whose indexes do not belong to $\mathcal{K}$, together with the rows whose indexes do not belong to $\underset{\sim}{\mathcal{K}}$, and we omit from $\mathbf{F}$ the columns whose indexes do not belong to $\underset{\sim}{\mathcal{K}}$. We denote the reduced matrices by $\mathbf{\Phi}_{\mathrm{r},(\alpha)}\in\mathbb{C}^{N_{\mathrm{p}}|\mathcal{K}|\times|\mathcal{K}|}$ and $\mathbf{F}_{\mathrm{r}}\in\mathbb{C}^{N_{\mathrm{p}}N_{\mathrm{CIR}}\times N_{\mathrm{p}}|\mathcal{K}|}$, respectively. Lastly, denote $\mathbf{H}_{\mathrm{r},(\alpha)}\triangleq\mathbf{S}\mathbf{F}_{\mathrm{r}}\mathbf{\Phi}_{\mathrm{r},(\alpha)}$, and $\mathbf{G}_{\mathrm{r},(\alpha)}\triangleq\left(\mathbf{H}^H_{\mathrm{r},(\alpha)}\mathbf{C}^{-1}_{\mathbf{w}}\mathbf{H}_{\mathrm{r},(\alpha)}\right)^{-1}$. Consequently, the observations can be approximated via

$$\mathbf{y}\simeq\mathbf{H}_{\mathrm{r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\mathrm{r}}+\mathbf{w},\qquad \mathbf{y}\overset{(approx.)}{\underset{\sim}{}}\mathcal{CN}\left(\mathbf{H}_{\mathrm{r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\mathrm{r}},\mathbf{C}_{\mathbf{w}}\right). \tag{12}$$

Note that (12) is an approximation as $\underset{\sim}{\mathbf{h}}$ is assumed to be only approximately sparse. Using the approximate sparse representation (12), the estimators in (7) can be reformulated as follows:

$$\hat{\alpha}^{\mathrm{CS}}=\underset{\alpha\in I_{1/2}}{\operatorname{argmax}}\left\{\mathbf{y}^H\mathbf{C}^{-1}_{\mathbf{w}}\mathbf{H}_{\mathrm{r},(\alpha)}\mathbf{G}_{\mathrm{r},(\alpha)}\mathbf{H}^H_{\mathrm{r},(\alpha)}\mathbf{C}^{-1}_{\mathbf{w}}\mathbf{y}\right\}, \tag{13a}$$

and

$$\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\hat{\alpha}^{\mathrm{CS}})}=\mathbf{G}_{\mathrm{r},(\hat{\alpha}^{\mathrm{CS}})}\mathbf{H}^H_{\mathrm{r},(\hat{\alpha}^{\mathrm{CS}})}\mathbf{C}^{-1}_{\mathbf{w}}\mathbf{y}. \tag{13b}$$

Note that a total of $|\mathcal{K}|+1$ parameters are estimated in (13), hence, the number of observations must satisfy $|\mathcal{K}|<N_{\mathrm{obs}}$. Since $|\mathcal{K}|$ is generally smaller than $N_{\mathrm{CIR}}=N_0L_{\mathrm{ch}}$, then, the estimator (13) requires a shorter PS compared to the PS length required for the JMLE via (7), and the spectral efficiency is accordingly improved. We emphasize that since the DDSF is only *approximately* sparse, the choice of the sparsity pattern impacts on the performance of the algorithm. Specifically, if the sparsity pattern $|\mathcal{K}|$ is too large, then the performance of the estimator may be degraded due to the increased number of estimated parameters. On the other hand, if $|\mathcal{K}|$ is too small, then dominant channel coefficients may be disregarded, thereby increasing the estimation error for the CIR coefficients. A possible approach to avoid having too many coefficients is to truncate the set $\mathcal{K}$ by setting an energy based threshold, e.g., fixing $0<\epsilon\ll 1$ and selecting $\mathcal{K}$ s.t. $\sum_{k\in\mathcal{K}}\sum_{p\in\mathcal{N}_{\mathrm{p}}}\left|\left[\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)\right]_{pN_{\mathrm{CIR}}+k}\right|^2\ge(1-\epsilon)\|\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)\|^2$.

Lastly, we note that the approximated ML estimator in (13b) is carried out under a misspecified observation model. Denote the true and approximated observations distributions of (5) and of (12) by $f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})$ and $f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\mathrm{r}})$, respectively, and by $\mathbf{B}_{\mathrm{r},(\alpha)}$ the projection matrix onto the column span of $\mathbf{C}^{-1/2}_{\mathbf{w}}\mathbf{H}_{\mathrm{r},(\alpha)}$. The estimate $\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\hat{\alpha}^{\mathrm{CS}})}$ satisfies the following property:

Lemma 3: The estimator $\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\hat{\alpha}^{\mathrm{CS}})}$ *is generally biased, yet, under the assumption that* α *is known,* $\mathbb{E}\left\{\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\alpha)}\right\}$ *minimizes the KLD between* $f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})$ *to* $f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\mathrm{r}})$ *w.r.t.* $\underset{\sim}{\mathbf{h}}_{\mathrm{r}}$*, specifically:*

$$\mathbb{E}\left\{\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\alpha)}\right\}=\underset{\sim}{\mathbf{h}}_{\mathrm{r}}+\mathbf{G}_{\mathrm{r},(\alpha)}\mathbf{H}^H_{\mathrm{r},(\alpha)}\mathbf{C}^{-1}_{\mathbf{w}}\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}-\mathbf{H}_{\mathrm{r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\mathrm{r}}\right). \tag{14a}$$

The resulting minimal KLD is equal to:

$$D\left(f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})||f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\mathrm{r}})\right)=\left\|\mathbf{B}^{\perp}_{\mathrm{r},(\alpha)}\mathbf{C}^{-1/2}_{\mathbf{w}}\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}\right\|^2_2. \tag{14b}$$

[A proof is given in Appendix B]

It follows that if $\forall k\notin\mathcal{K}$, $[\underset{\sim}{\mathbf{h}}]_k=0$, implying that $\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}=\mathbf{H}_{\mathrm{r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\mathrm{r}}$, then, from (14a), $\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\alpha)}$ is an unbiased estimator of $\underset{\sim}{\mathbf{h}}_{\mathrm{r}}$. Accordingly, in such circumstances $f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})$ is equal to $f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\mathrm{r}})$, which due to Lemma 3 results in a minimal KLD, which is equal to zero. In practice, $\underset{\sim}{\mathbf{h}}_{\mathrm{r}}$ is estimated via (13b) using the CFO estimate $\hat{\alpha}^{\mathrm{CS}}$ which is a random variable dependent on $\mathbf{y}$ via (13a). Due to this dependency, obtaining an explicit expression for $\mathbb{E}\left\{\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\hat{\alpha}^{\mathrm{CS}})}\right\}$ is intractable, however, at sufficiently high signal-to-noise ratio (SNR) such that $\hat{\alpha}^{\mathrm{CS}}\simeq\alpha$ and if $\forall k\notin\mathcal{K}$, $[\underset{\sim}{\mathbf{h}}]_k=0$, then $\hat{\underset{\sim}{\mathbf{h}}}^{\mathrm{CS}}_{\mathrm{r},(\hat{\alpha}^{\mathrm{CS}})}$ yields an unbiased estimate of $\underset{\sim}{\mathbf{h}}_{\mathrm{r}}$.

Complexity Analysis: Computing the joint estimator via (13) is done similarly to computing the joint estimator via (7). Therefore, the computational complexity of (13) is the same as that of (7) except for the following two differences:

1) The number of estimated CIR coefficients is decreased from $N_{\mathrm{CIR}}=N_0L_{\mathrm{ch}}$ to $|\mathcal{K}|<N_{\mathrm{CIR}}$. This decreases the computational complexity proportionally: As the complexity of (7) is $\mathcal{O}\left(T\cdot N^3_{\mathrm{CIR}}\right)$, we conclude that the complexity of (13) is $\mathcal{O}\left(T\cdot|\mathcal{K}|^3\right)$.
2) Finding the BPDN solution $\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)$ introduces additional computational complexity which depends on the details of the algorithm applied. For example, group sparsity estimation can be implemented as in [33, Sec. 4.3], with a computational complexity on the order of $\mathcal{O}(N_{\mathrm{p}}N_{\mathrm{CIR}})$

complex multiplications and $\mathcal{O}(N_{\mathrm{p}}N_{\mathrm{CIR}})$ complex additions. It follows that the complexity of the sparsity pattern estimation is negligible compared to $\mathcal{O}\left(T\cdot|\mathcal{K}|^3\right)$.

To summarize, the overall computational complexity of (13) is on the order of $\mathcal{O}\left(T\cdot|\mathcal{K}|^3\right)$.

B. Low Complexity Implementation of the JMLE Based on Sparse DDSF Approximation

While the joint ML estimator based on sparse DDSF approximation (13) improves the spectral efficiency compared to the JMLE in (7), evaluating (13a) still requires a computational complexity on the order of $\mathcal{O}\left(T\cdot|\mathcal{K}|^3\right)$. In order to further reduce the computational complexity, we now propose a sub-optimal estimator based on the following rationale: We divide the observations vector $\mathbf{y}$ into N_{p} segments of length N_0, denoted $\{\mathbf{y}_p\}_{p=0}^{N_{\mathrm{p}}-1}$. Let $\left\{\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}\right\}_{p=0}^{N_{\mathrm{p}}-1}$ denote the phase shifted version of DDSF coefficient at each segment $p\in\mathcal{N}_{\mathrm{p}}$, $\underset{\sim}{\mathbf{h}}_{\mathrm{r},p}^{(\alpha)}\triangleq\underset{\sim}{\mathbf{h}}_{\mathrm{r}}e^{j2\pi\alpha p}$. We then estimate $\underset{\sim}{\mathbf{h}}_{\mathrm{r},p}^{(\alpha)}$ at each segment p based only on $\mathbf{y}_p$, and denote the set of estimates obtained with $\left\{\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}\right\}_{p=0}^{N_{\mathrm{p}}-1}$. Lastly, we use the estimates $\left\{\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}\right\}_{p=0}^{N_{\mathrm{p}}-1}$ to obtain estimates for the DDSF coefficients and for the CFO. We note that this approach is clearly sub-optimal, as the separate estimates do not utilize the correlation across sequences of length N_0, however, the resulting estimator has a much lower computational complexity compared to the estimator based on (13). Since at each segment $p\in\mathcal{N}_{\mathrm{p}}$, the DDSF $\underset{\sim}{\mathbf{h}}_{\mathrm{r},p}^{(\alpha)}$ is of dimension $|\mathcal{K}|$ while the observations $\mathbf{y}_p$ is of dimension N_0, then this approach is applicable only when the cardinality of the sparsity pattern satisfies $|\mathcal{K}|\leq N_0$.[2]

Specifically, let $\mathbf{S}_p\in\mathbb{C}^{N_0\times N_{\mathrm{CIR}}}$ be a sub-matrix of $\mathbf{S}$ consisting of the entries $[\mathbf{S}]_{n,l}$ with indexes $pN_0\leq n<(p+1)N_0$ and $pN_{\mathrm{CIR}}\leq l<(p+1)N_{\mathrm{CIR}}$, and set $\underline{\mathbf{F}}_{\mathrm{r}}\in\mathbb{C}^{N_{\mathrm{CIR}}\times|\mathcal{K}|}$ to be a matrix whose columns are the columns of the matrix $\mathbf{F}_{N_0}^{-1}\otimes\mathbf{I}_{L_{\mathrm{ch}}}$ corresponding to the indexes of the set $\mathcal{K}$. Lastly, define the vectors $\mathbf{y}_p,\mathbf{w}_p\in\mathbb{C}^{N_0}$ s.t. $\left[\mathbf{y}_p\right]_n=\left[\mathbf{y}\right]_{pN_0+n}$ and $\left[\mathbf{w}_p\right]_n=[\mathbf{w}]_{pN_0+n}$, $p\in\mathcal{N}_{\mathrm{p}}$, $n\in\mathcal{N}_0$. As $w[n]$ is WSCS, the covariance matrix of $\mathbf{w}_p$ *does not depend on* p, and we thus denote it by $\underline{\mathbf{C}}_{\mathbf{w}}\triangleq\mathbb{E}\left\{\mathbf{w}_p\mathbf{w}_p^H\right\}$. It follows from (12) that

$$\mathbf{y}_p\simeq\mathbf{S}_p\underline{\mathbf{F}}_{\mathrm{r}}\underset{\sim}{\mathbf{h}}_{\mathrm{r},p}^{(\alpha)}+\mathbf{w}_p,\quad \mathbf{y}_p\overset{(approx.)}{\sim}\mathcal{CN}\left(\mathbf{S}_p\underline{\mathbf{F}}_{\mathrm{r}}\underset{\sim}{\mathbf{h}}_{\mathrm{r},p}^{(\alpha)},\underline{\mathbf{C}}_{\mathbf{w}}\right). \quad (15)$$

Define $\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}\triangleq\left(\underline{\mathbf{F}}_{\mathrm{r}}^H\mathbf{S}_p^H\underline{\mathbf{C}}_{\mathbf{w}}^{-1}\mathbf{S}_p\underline{\mathbf{F}}_{\mathrm{r}}\right)^{-1}$. Since for any given value of α and for any $p\in\mathcal{N}_{\mathrm{p}}$, (15) corresponds to a linear relationship between $\mathbf{y}_p$ and $\underset{\sim}{\mathbf{h}}_{\mathrm{r},p}^{(\alpha)}$ with additive Gaussian noise, it follows that *the efficient estimator* of $\underset{\sim}{\mathbf{h}}_{\mathrm{r},p}^{(\alpha)}$ from $\mathbf{y}_p$ is given by [32, Ch. 4.5]:

$$\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}=\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}\underline{\mathbf{F}}_{\mathrm{r}}^H\mathbf{S}_p^H\underline{\mathbf{C}}_{\mathbf{w}}^{-1}\mathbf{y}_p,\ \ \hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}\sim\mathcal{CN}\left(\underset{\sim}{\mathbf{h}}_{\mathrm{r}}e^{j2\pi\alpha p},\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}\right). \quad (16)$$

Ignoring the correlation between the estimates $\left\{\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}\right\}_{p=0}^{N_{\mathrm{p}}-1}$, we obtain the reduced complexity CS-based (RCS) joint estimate of $\underset{\sim}{\mathbf{h}}_{\mathrm{r}}$ and α as the solution to the following WLS problem:

$$\left(\hat{\alpha}^{\mathrm{RCS}},\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r}}^{\mathrm{RCS}}\right)=\underset{\substack{\alpha\in I_{1/2},\\ \underset{\sim}{\mathbf{h}}_{\mathrm{r}}\in\mathbb{C}^{|\mathcal{K}|}}}{\operatorname{argmin}}\left\{\sum_{p=0}^{N_{\mathrm{p}}-1}\left(\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}-e^{j2\pi\alpha p}\underset{\sim}{\mathbf{h}}_{\mathrm{r}}\right)^H\cdot\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}^{-1}\left(\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}-e^{j2\pi\alpha p}\underset{\sim}{\mathbf{h}}_{\mathrm{r}}\right)\right\}. \quad (17)$$

To implement the reduced complexity estimator, define the $|\mathcal{K}|\times|\mathcal{K}|$ matrix $\mathbf{C}_{\mathrm{inv}}\triangleq\left(\sum_{p=0}^{N_{\mathrm{p}}-1}\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}^{-1}\right)^{-1}$, and the function $\tilde{\mathbf{x}}(\cdot):I_{1/2}\mapsto\mathbb{C}^{|\mathcal{K}|}$, such that $\tilde{\mathbf{x}}(\alpha)\triangleq\sum_{p=0}^{N_{\mathrm{p}}-1}\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}^{-1}\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}e^{-j2\pi\alpha p}$. In Appendix C we show that the reduced complexity estimator (17) can be obtained via:

$$\hat{\alpha}^{\mathrm{RCS}}=\underset{\alpha\in I_{1/2}}{\operatorname{argmax}}\ \left(\tilde{\mathbf{x}}(\alpha)\right)^H\mathbf{C}_{\mathrm{inv}}\tilde{\mathbf{x}}(\alpha), \quad (18a)$$

and

$$\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},(\hat{\alpha}^{\mathrm{RCS}})}^{\mathrm{RCS}}=\mathbf{C}_{\mathrm{inv}}\tilde{\mathbf{x}}(\hat{\alpha}^{\mathrm{RCS}}). \quad (18b)$$

Lastly, we write $\mathbf{C}_{\mathrm{inv}}$ and $\tilde{\mathbf{x}}(\alpha)$ using the scenario parameters: Substituting $\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}$ into the definition of $\mathbf{C}_{\mathrm{inv}}$ we obtain

$$\mathbf{C}_{\mathrm{inv}}=\left(\sum_{p=0}^{N_{\mathrm{p}}-1}\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}^{-1}\right)^{-1}=\left(\sum_{p=0}^{N_{\mathrm{p}}-1}\underline{\mathbf{F}}_{\mathrm{r}}^H\mathbf{S}_p^H\underline{\mathbf{C}}_{\mathbf{w}}^{-1}\mathbf{S}_p\underline{\mathbf{F}}_{\mathrm{r}}\right)^{-1}. \quad (19a)$$

Then, substituting $\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}$ and (16) into the definition of $\tilde{\mathbf{x}}(\alpha)$ we obtain

$$\begin{aligned}\tilde{\mathbf{x}}(\alpha)&=\sum_{p=0}^{N_{\mathrm{p}}-1}\underset{\sim}{\mathbf{C}}_{\mathrm{r},p}^{-1}\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r},p}^{(\alpha)}e^{-j2\pi\alpha p}\\&=\sum_{p=0}^{N_{\mathrm{p}}-1}\underline{\mathbf{F}}_{\mathrm{r}}^H\mathbf{S}_p^H\underline{\mathbf{C}}_{\mathbf{w}}^{-1}\mathbf{y}_pe^{-j2\pi\alpha p}.\end{aligned} \quad (19b)$$

Complexity Analysis: The complexity analysis in terms of complex multiplications and complex additions is detailed in Table II. In the following we provide a brief outline of the derivations of the expressions in Table II. As in the complexity analysis of the JMLE, α is estimated by evaluating Eqn. (18a) at T test points, where T is significantly larger than N_{CIR}. Accordingly, (19b) should also be evaluated T times. The expressions in (20a)–(20d) correspond to calculating intermediate quantities in the evaluation of (19a); the expressions in (20e)–(20g) correspond to calculating intermediate quantities in the evaluation of (19b); the expression (20h) corresponds to (18a) hence, it is evaluated T times; and the expression (20i) corresponds to (18b), hence, it is evaluated only once. Note that the complexities of (20b), (20d), and (20f) follow from the sparse structure of the matrices $\mathbf{S}_p$ and $\underline{\mathbf{F}}_{\mathrm{r}}$. Lastly, the term (20j) represents the computational complexity required for obtaining the BPDN solution $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}^{o}(\alpha)$ via the algorithm described in [33, Sec. 4.3]. Note that the estimator (18) does not require a matrix inversion at each of the T test points, and its computational complexity is dominated by the

[2]Note that when $|\mathcal{K}|>N_0$, we can always decrease $|\mathcal{K}|$ by taking only the N_0 coefficients with the largest magnitudes out of the $|\mathcal{K}|$ coefficients produced by the sparsity estimation process. This truncation clearly increases the estimation error.

TABLE II
COMPUTATIONAL COMPLEXITY OF (18)

	Term	Complex $\times$	Complex $+$
(20*a*)	$\underline{\mathbf{C}}_{\mathbf{w}}^{-1}$	$\mathcal{O}\left(N_0^3\right)$	$\mathcal{O}\left(N_0^3\right)$
(20*b*)	$\mathbf{S}_p\underline{\mathbf{F}}_{\mathrm{r}}$	$N_{\mathrm{p}}N_0\lvert\mathcal{K}\rvert$	
(20*c*)	$\sum_{p=0}^{N_{\mathrm{p}}-1}\underline{\mathbf{F}}_{\mathrm{r}}^H\mathbf{S}_p^H\underline{\mathbf{C}}_{\mathbf{w}}^{-1}\mathbf{S}_p\underline{\mathbf{F}}_{\mathrm{r}}$	$N_{\mathrm{p}}\lvert\mathcal{K}\rvert N_0\left(N_0+\frac{\lvert\mathcal{K}\rvert+1}{2}\right)$	$\lvert\mathcal{K}\rvert\left(N_{\mathrm{p}}(N_0-1)(N_0+\frac{\lvert\mathcal{K}\rvert+1}{2})+\frac{\lvert\mathcal{K}\rvert+1}{2}(N_{\mathrm{p}}-1)\right)$
(20*d*)	$\mathbf{C}_{\mathrm{inv}}$	$\mathcal{O}\left(\lvert\mathcal{K}\rvert^3\right)$	$\mathcal{O}\left(\lvert\mathcal{K}\rvert^3\right)$
(20*e*)	$\underline{\mathbf{C}}_{\mathbf{w}}^{-1}\mathbf{y}_p$	$N_{\mathrm{p}}N_0^2$	$N_{\mathrm{p}}N_0(N_0-1)$
(20*f*)	$(\mathbf{S}_p\underline{\mathbf{F}}_{\mathrm{r}})^H(\underline{\mathbf{C}}_{\mathbf{w}}^{-1}\mathbf{y}_p)$	$N_{\mathrm{p}}\lvert\mathcal{K}\rvert N_0$	$N_{\mathrm{p}}\lvert\mathcal{K}\rvert(N_0-1)$
(20*g*)	$\tilde{\mathbf{x}}(\alpha)$	$\lvert\mathcal{K}\rvert\cdot\mathcal{O}(T\log(N_{\mathrm{p}}))$	$\lvert\mathcal{K}\rvert\cdot\mathcal{O}(T\log(N_{\mathrm{p}}))$
(20*h*)	$\tilde{\mathbf{x}}(\alpha)^H\mathbf{C}_{\mathrm{inv}}\tilde{\mathbf{x}}(\alpha)$	$2T\cdot\lvert\mathcal{K}\rvert^2$	$T\left(\lvert\mathcal{K}\rvert^2-1\right)$
(20*i*)	$\mathbf{C}_{\mathrm{inv}}\tilde{\mathbf{x}}(\hat{\alpha}^{\mathrm{RCS}})$	$\lvert\mathcal{K}\rvert^2$	$\lvert\mathcal{K}\rvert(\lvert\mathcal{K}\rvert-1)$
(20*j*)	$\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}^o(\alpha)$	$\mathcal{O}(N_{\mathrm{p}}N_{\mathrm{CIR}})$	$\mathcal{O}(N_{\mathrm{p}}N_{\mathrm{CIR}})$

computations of (20g)-(20h). We conclude that implementing (18) has a computational complexity of an order of $\mathcal{O}(T\lvert\mathcal{K}\rvert^2)$, which is considerably lower compared to the JMLE (7) and to the CS–based estimator of (13).

IV. PILOT SEQUENCE DESIGN FOR SPARSE LPTV CHANNELS

In this section we suggest a method for designing the PS, $\left\{s\left[n\right]\right\}_{n=-L_{\mathrm{ch}}+1}^{N_{\mathrm{obs}}}$ to achieve good estimation performance. A common approach for designing PSs is to minimize the trace of the CRB for the estimation of the CIR coefficients, which optimizes the performance of the JMLE, as its performance approaches to the CRB when the number of observations and the SNR are sufficiently high [17]. This approach was applied in [17] and [20], which assumed that the noise is AWGN. In such cases, the CRB for the CIR coefficients and the CFO does not depend on the CFO. When the noise is ACGN, the CRB *depends on both the unknown CIR and on the CFO*, (the CRB is detailed in Appendix D) as well as on the autocorrelation of the noise. This dependence on the unknown parameters renders it impossible to implement the PS design methods proposed in [17] and [20].[3] Furthermore, our CS-based estimators may be implemented when the number of observations is smaller than the number of unknown parameters, i.e., $N_{\mathrm{p}}\leq L_{\mathrm{ch}}$, and for such scenarios the CRB does not exist [32, Ch. 7.8]. To circumvent the above difficulties, we propose an alternative approach for designing the PS, which aims at improving the performance of the sparsity pattern estimation, thus improving the performance of the CS–based estimation schemes, as detailed in Lemma 3.

[3]We note that [17] proposed a design method which constrained the PS to obtain a CRB for the estimation of the CFO which depends only on the magnitude of the CIR coefficients vector, i.e., on $\|\mathbf{h}\|$. This simplification comes at the cost of increasing the CRB compared to the minimal achievable CRB for a given channel realization, when the observations are corrupted by AWGN. However, adapting this approach to correlated noise results in an optimization function which still depends on the CFO.

Recall first that in order to estimate the sparsity pattern, the input-output relationship was stated as a CS problem in (10), where $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ corresponds to the unknown group-sparse vector and the product $\mathbf{SF}\in\mathbb{C}^{N_{\mathrm{obs}}\times N_{\mathrm{p}}N_{\mathrm{CIR}}}$ corresponds to the *sensing matrix* [22, Ch. 1.4]. The performance of sparsity pattern estimation algorithms in correctly identifying the dominant coefficients of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ largely depends on the *mutual coherence (MC)* of the sensing matrix [22], [23], defined as the maximal absolute value of the inner products between all pairs of different normalized columns of the sensing matrix [22, Definition 1.5]. More specifically, for group-sparse signals, the relevant coherence metrics are the *sub-coherence* and the *group coherence* [35, Sec. II-B]. In the following we propose to design the PS such that these metrics are minimized, thus improving the performance of the group-sparsity pattern estimation.

Define $\mathbf{A}\triangleq\mathbf{SF}$, and note that the entry of $\mathbf{A}$ at row $r=p_1N_0+n$ and column $c=p_2N_{\mathrm{CIR}}+kL_{\mathrm{ch}}+l$, for $p_1,p_2\in\mathcal{N}_{\mathrm{p}}$, $n,k\in\mathcal{N}_0$, $l\in\mathcal{L}_{\mathrm{ch}}$, is given by $[\mathbf{A}]_{r,c}=\frac{1}{\sqrt{N_0}}s\left[p_1N_0+n-l\right]e^{j\frac{2\pi}{N_0}kn}\delta[p_1-p_2]$. It thus follows that $\mathbf{A}$ is a block-diagonal matrix consisting of N_{p} blocks, each of size $N_0\times N_{\mathrm{CIR}}$. Recall that the sparsity pattern estimation procedure aims at finding a group sparse vector $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$, in which the coefficients at indexes $\left\{[\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)]_{pN_{\mathrm{CIR}}+kL_{\mathrm{ch}}+l}\right\}_{p\in\mathcal{N}_{\mathrm{p}}}$ belong to the same group $\forall k\in\mathcal{N}_0,l\in\mathcal{L}_{\mathrm{ch}}$. Consequently, each group of coefficients from $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ is associated with an $N_{\mathrm{obs}}\times N_{\mathrm{p}}$ matrix consisting of a single column from each block of $\mathbf{A}$. Let $\mathbf{A}[k,l]$ be the $N_{\mathrm{p}}\cdot N_0\times N_{\mathrm{p}}$ *normalized* matrix associated with the (k,l) group, $k\in\mathcal{N}_0$, $l\in\mathcal{L}_{\mathrm{ch}}$, i.e., the column at index $p\in\mathcal{N}_{\mathrm{p}}$ of $\mathbf{A}[k,l]$ is obtained as the column at index $pN_{\mathrm{CIR}}+kL_{\mathrm{ch}}+l$ of the matrix $\mathbf{A}$, normalized to have a unit norm. We can now define the coherence measures for our group-sparse setup:

The *sub-coherence* of $\mathbf{A}$ is defined as the maximal MC among all sub-matrices of $\mathbf{A}$, i.e., the maximal MC among all sub-matrices $\mathbf{A}[k,l]$, $k\in\mathcal{N}_0$, $l\in\mathcal{L}_{\mathrm{ch}}$ [35, eq. (8)]: Letting

$\mathrm{MC}(\mathbf{X})$ denote the MC of a matrix $\mathbf{X}$, the sub-coherence of $\mathbf{A}$ is defined as:

$$\mathrm{SC}(\mathbf{A}) = \max_{k\in\mathcal{N}_0, l\in\mathcal{L}_{\mathrm{ch}}} \left\{\mathrm{MC}(\mathbf{A}[k,l])\right\}. \tag{21}$$

As each column of $\mathbf{A}[k,l]$ is taken from a different sub-block of $\mathbf{A}$, the columns of $\mathbf{A}[k,l]$ are *orthogonal* for any $(k,l)\in\mathcal{N}_0\times\mathcal{L}_{\mathrm{ch}}$, hence, $\mathrm{MC}(\mathbf{A}[k,l])=0$, $\forall k\in\mathcal{N}_0$, $\forall l\in\mathcal{L}_{\mathrm{ch}}$. Consequently, $\mathrm{SC}(\mathbf{A})=0$ for any PS, and this measure is not viable for PS design. Note that $\mathrm{SC}(\mathbf{A})=0$ does not imply that $\mathrm{MC}(\mathbf{A})=0$, as the columns of *different* sub-matrices, e.g., of $\mathbf{A}[k_1,l_1]$ and of $\mathbf{A}[k_2,l_2]$ for $(k_1,l_1)\neq(k_2,l_2)$, are not necessarily orthogonal.

The *group coherence* of $\mathbf{A}$, denoted $\mathrm{GC}(\mathbf{A})$, is defined as [35, eq. (6)]:

$$\mathrm{GC}(\mathbf{A}) \triangleq \frac{1}{N_{\mathrm{p}}} \max_{\substack{k_1,k_2\in\mathcal{N}_0, l_1,l_2\in L_{ch}\\ (k_1,l_1)\neq(k_2,l_2)}} \rho\left\{\left(\mathbf{A}[k_1,l_1]\right)^H\mathbf{A}[k_2,l_2]\right\}.$$

Note that the $N_{\mathrm{p}}\times N_{\mathrm{p}}$ matrix $\left(\mathbf{A}[k_1,l_1]\right)^H\mathbf{A}[k_2,l_2]$ is diagonal $\forall k_1,k_2\in\mathcal{N}_0$, $\forall l_1,l_2\in\mathcal{L}_{\mathrm{ch}}$, and that its diagonal entries are given by:

$$\left[\left(\mathbf{A}[k_1,l_1]\right)^H\mathbf{A}[k_2,l_2]\right]_{p,p} = \sum_{n=0}^{N_0-1} \frac{s\left[pN_0+n-l_1\right]}{\sqrt{\sum_{n=0}^{N_0-1}\left|s\left[pN_0+n-l_1\right]\right|^2}} \cdot \frac{s^*[pN_0+n-l_2]e^{j\frac{2\pi}{N_0}n(k_1-k_2)}}{\sqrt{\sum_{n=0}^{N_0-1}\left|s\left[pN_0+n-l_2\right]\right|^2}}, \tag{22}$$

$\forall p\in\mathcal{N}_{\mathrm{p}}$. Since $\left(\mathbf{A}[k_1,l_1]\right)^H\mathbf{A}[k_2,l_2]$ is diagonal, its singular values are the absolute values of its diagonal entries [36, p. 555]. Furthermore, Eqn. (22) is equal to the inner product between two normalized columns of $\mathbf{A}$: The column at index $pN_{\mathrm{CIR}}+k_1L_{\mathrm{ch}}+l_1$ and the column at index $pN_{\mathrm{CIR}}+k_2L_{\mathrm{ch}}+l_2$. Consequently, the maximal absolute value of (22) is equal to $\mathrm{MC}(\mathbf{A})$. We thus obtain that $\mathrm{GC}(\mathbf{A})$ is equal to $\mathrm{MC}(\mathbf{A})$, up to a factor of N_{p}^{-1}.

We next design the PS such that the group coherence $\mathrm{GC}(\mathbf{A})$ is minimized. As in [17], to reduce the feasible region, we restrict the PS to satisfy a constant unit magnitude, i.e., $|s[n]|=1$, $-L_{\mathrm{ch}}<n<N_{\mathrm{obs}}$. This constraint also guarantees that the columns of $\mathbf{A}$ all have a unit norm, which further simplifies the PS design procedure. Additionally, we note that the matrices $\mathbf{H}_{(\alpha)}$ and $\mathbf{H}_{\mathrm{r},(\alpha)}$ must both be of full rank in order to allow the implementation of the estimators in (7) and in (13) [32, Th. 4.1]. From (4) we note that α affects the rows of $\mathbf{H}_{(\alpha)}$ and of $\mathbf{H}_{\mathrm{r},(\alpha)}$ via a multiplication by a complex scalar, hence, the ranks of $\mathbf{H}_{(\alpha)}$ and $\mathbf{H}_{\mathrm{r},(\alpha)}$ *do not depend on* α. Furthermore, note that the product $\mathbf{F}_{\mathrm{r}}\mathbf{\Phi}_{\mathrm{r},(\alpha)}$ can be obtained from $[1, e^{j2\pi\alpha}..., e^{j2\pi\alpha(N_{\mathrm{p}}-1)}]^T\otimes\mathbf{F}_{N_0}^{-1}\otimes\mathbf{I}_{L_{\mathrm{ch}}}$, after removing the columns whose indexes are not contained within $\mathcal{K}$. Thus, it follows that columns of $\mathbf{F}_{\mathrm{r}}\mathbf{\Phi}_{\mathrm{r},(\alpha)}$ are orthogonal for any $\mathcal{K}\subseteq\mathcal{N}_{\mathrm{CIR}}$, while the PS entries affect only the entries of the matrix $\mathbf{S}$, which does not depend on $\mathcal{K}$. It is therefore concluded that the rank of $\mathbf{H}_{\mathrm{r},(\alpha)}$ depends only the PS entries and on $|\mathcal{K}|$, and not on the specific indexes contained in $\mathcal{K}$.

From the above observation we conclude that the PS design which minimizes $\mathrm{GC}(\mathbf{A})$ can be obtained via the following optimization problem:

$$\min_{\{s[n]\}_{n=-L_{\mathrm{ch}}+1}^{N_{\mathrm{obs}}}} \left\{ \max_{\substack{k_1,k_2\in\mathcal{N}_0, l_1,l_2\in\mathcal{L}_{\mathrm{ch}},\\ (k_1,l_1)\neq(k_2,l_2), p\in\mathcal{N}_{\mathrm{p}}}} \left\{ \left| \sum_{n=0}^{N_0-1} s\left[pN_0+n-l_1\right] \cdot s^*[pN_0+n-l_2]e^{j\frac{2\pi}{N_0}n(k_1-k_2)} \right| \right\} \right\} \tag{23a}$$

$$\text{s.t.} \quad \underset{-L_{\mathrm{ch}}<n<N_{\mathrm{obs}}}{|s[n]|=1} \;;\; \mathrm{rank}\left(\mathbf{H}_{(\alpha)}\right)=\min\left\{N_{\mathrm{CIR}}, N_{\mathrm{obs}}\right\}. \tag{23b}$$

We note that it looks intuitively appealing to consider periodic PSs with a cyclic-prefix structure, i.e., considering PSs which satisfy $s[n]=s[n+pN_0]$, for all $n\in\mathcal{N}_0$, $p\in\mathcal{N}_{\mathrm{p}}$ and $s[-l]=s[N_0-l]$, for all $l\in\mathcal{L}_{\mathrm{ch}}$, since under such constraints (23a) is independent of p. A similar PS structure was used in [19] to facilitate CFO estimation without estimating the CIR. However, periodic PSs result in the matrices $\mathbf{H}_{(\alpha)}$ and $\mathbf{H}_{\mathrm{r},(\alpha)}$ having dependent rows due to (4). Thus, periodic PSs violate the full rank constraint on $\mathbf{H}_{(\alpha)}$ and $\mathbf{H}_{\mathrm{r},(\alpha)}$, hence, *the optimal PS is not periodic although the channel is periodic.* Consequently, the CFO estimation scheme detailed in [19] cannot be extended to incorporate CIR estimation.

As an analytic solution to (23a) is difficult to obtain, a possible approach to obtain a PS with good sparsity estimation characteristics is as follows:

1) Randomly select a set of initial PSs with elements uniformly distributed on the unit circle.
2) For each randomly selected PS, perform numerical optimization of (23a) while ignoring the rank constraint to achieve convergences to a possibly local minima.
3) For each locally optimal PS, verify that the rank constraint is satisfied. If the rank constraint is not satisfied, discard the resulting PS.
4) Out of the remaining PSs with full rank $\mathbf{H}_{(\alpha)}$, choose the PS with the smallest $\mathrm{GC}(\mathbf{A})$.

V. Numerical Examples and Discussion

In this section we compare the performance of the JMLE detailed in Subsection II-C, and the sparsity-based estimators proposed in Subsections III-A and III-B, referred to henceforth as CS-AML and RCS-AML, respectively, where AML stands for "approximated JMLE", via a simulations study. We consider three scenarios: Two scenarios corresponding to *common practical* periodic channels with an approximately sparse DDSF: An NB-PLC channel and a mobile radio channel modeled via exponential basis functions. The third scenario is a scenario in which the DDSF is exactly sparse. The parameters of the simulated channels are as follows:

1) *NB-PLC channel* - The NB-PLC CIR $g[n,l]$ is generated as detailed in [37, Sec. V] with a period of $N_0=20$ and a memory of $L_{\mathrm{ch}}=5$; the CFO is set to $\alpha=0.2$. The noise $w[n]$ in the NB-PLC simulations is generated according to the WSCS noise model detailed in [1] and in the IEEE P1901.2 standard [38], with a set of typical parameters based on [38, Appendix G, LV11].

2) *Mobile radio channel* - The mobile radio CIR $g[n,l]$ is generated as detailed in [9, Sec. 6] with parameters $T_1 = 0$, $T_2 = 4$, $T_3 = 5$, $\theta_{l,1} = 1$, $\theta_{l,2} = j \cdot (-0.5j)^l$ and $\theta_{l,3} = 2 \cdot (-0.5j)^l$, for $l \in \mathcal{L}_{\text{ch}}$. This selection corresponds to the parameters in [10, Sec. VI], scaled to achieve a period of $N_0 = 20$ and a memory of $L_{\text{ch}} = 5$. Following [27], the CFO value considered is $\alpha = 0.1$. The noise $w[n]$ is a stationary autoregressive process with poles at $0.8 \pm 0.2j$, as in [39, Sec. VI].[4]
3) *Exactly sparse channel* - In order to demonstrate the effectiveness of the proposed CS-based estimators, we also tested the proposed estimators in a scenario in which the DDSF is *exactly sparse*. This scenario is based on the mobile radio scenario described in item 2 above, where in order to achieve exact sparsity all coefficients in the vector $\underset{\sim}{\mathbf{h}}$, except for the nine coefficients with the largest magnitudes, were set to zero.

In all simulations reported in this section, sparsity pattern estimation was done using the BPDN algorithm detailed in [33] and [34] with $\sigma^2 = 1.05 \cdot \mathbb{E}\left\{\|\mathbf{w}\|^2\right\}$. The sparsity pattern $\mathcal{K}$ is obtained from the output of the BPDN algorithm, $\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)$, by taking the minimal number of indexes which accounts for 99.5% of the energy of $\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)$, i.e., $\mathcal{K}$ is the set with the smallest cardinality which satisfies $\sum_{k\in\mathcal{K}}\sum_{p\in\mathcal{N}_{\text{p}}} |\left[\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)\right]_{pN_{\text{CIR}}+k}|^2 \geq 0.995 \cdot \|\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)\|^2$. For the RCS-AML estimator (18), $|\mathcal{K}|$ cannot be larger than N_0. Hence, for this estimator, if the sparsity pattern $\mathcal{K}$ estimated using the above energy criterion has $|\mathcal{K}| > N_0$, then the set $\mathcal{K}$ is truncated to contain only the indexes corresponding to the N_0 groups of entries of $\underset{\sim}{\mathbf{h}}^o_{\mathbf{\Phi}}(\alpha)$ with the largest magnitudes. Estimation performance is reported in the following for two lengths of observed sequences corresponding to $N_{\text{p}} = 4$ and $N_{\text{p}} = 6$. We note that for $N_{\text{p}} = 4$, the number of observations, $N_{\text{obs}} = N_{\text{p}} \cdot N_0$, is smaller than the total number of parameters $N_0 \cdot L_{\text{ch}} + 1$, and consequently, the JMLE (7) cannot be implemented.

The PS $\{s[n]\}_{n=-L_{\text{ch}}+1}^{N_{\text{obs}}}$ used in the simulations was designed using the method detailed in Section IV, optimized over 100 randomly generated initial PSs. The estimation of α was implemented using a grid search over $T = 10^4$ equally spaced points in the interval $I_{1/2}$. The SNR for the simulations is defined as[5] $\text{SNR} \triangleq \|\mathbf{y} - \mathbf{w}\|_2^2 / \mathbb{E}\left\{\|\mathbf{w}\|_2^2\right\}$.

In the following, the CIR estimation performance is evaluated using the normalized MSE (NMSE), defined as $\mathbb{E}\left\{\|\underset{\sim}{\hat{\mathbf{h}}} - \underset{\sim}{\mathbf{h}}\|^2\right\} / \|\underset{\sim}{\mathbf{h}}\|^2$. Since the parameter α is periodic, its estimation performance is measured via the mean-squared periodic error (MSPE), defined as $\text{MSPE}(\hat{\alpha}) \triangleq \mathbb{E}\left\{\langle \hat{\alpha} - \alpha \rangle^2\right\}$ [40, Sec. II]. Lastly, the performance of sparsity pattern estimation is evaluated in terms of the mean ratio between the CIR energy contained in $\mathcal{K}$ and the total energy of the CIR, evaluated using[6] $r(\mathcal{K}) \triangleq \mathbb{E}_{\mathcal{K}}\left\{\sum_{k\in\mathcal{K}} |[\underset{\sim}{\mathbf{h}}]_k|^2\right\} / \|\underset{\sim}{\mathbf{h}}\|^2$. Note that the term $1 - r(\mathcal{K})$ constitutes a lower bound on the NMSE achievable for the estimated CIR. Note also that small mean ratio values imply that a significant part of the estimated DDSF coefficients is set to zero and is thus not estimated by the CS-based estimators. Consequently, when the mean ratio is low, a significant bias should be expected in the estimated DDSF coefficients.

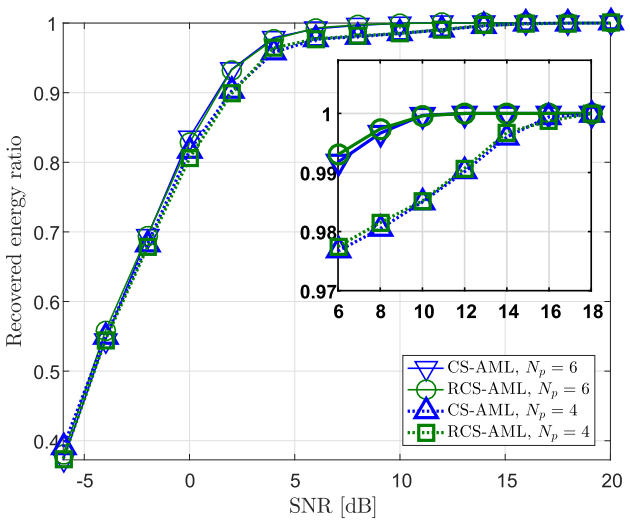


Fig. 1. Performance of sparsity pattern estimation, sparse mobile radio scenario.

A. Performance Evaluation in the Exactly Sparse Scenario

The sparsity pattern estimation performance is depicted in Fig. 1. As can be observed in the figure, since $N_0 > 9$ then, at high SNRs both the CS-AML and RCS-AML correctly identify *all* of the non-zero coefficients of $\underset{\sim}{\mathbf{h}}$. Consequently, we confirm that the sparsity estimation algorithm presented in Subsection III-A is able to recover the sparsity pattern of an exactly sparse DDSF.

The CIR NMSE estimation performance and the CFO MSPE estimation performance are depicted in Figs. 2 and 3, respectively. The dashed curves depict the CRB evaluated without knowledge of the sparsity pattern for $N_{\text{p}} = 6$ (the detailed derivation of the CRB is provided in Appendix D). Note that for $N_{\text{p}} = 4 < L_{\text{ch}}$, the number of observations N_{obs} is smaller than the number of unknown parameters, and thus the CRB does not exist [32, Ch. 7.8]. In addition, the dash-dot curves depict the "oracle CRBs" for the exactly sparse scenario, which are the CRBs for the joint estimation of α and $\underset{\sim}{\mathbf{h}}_{\text{r}}$, evaluated subject to having exact knowledge of the sparsity pattern, as in [28, Sec. 5]. The oracle CRBs exist for both $N_{\text{p}} = 4$ and $N_{\text{p}} = 6$. Note that at high SNR values, the CRB also lower bounds the MSPE for the estimation of α, since at high SNRs, the MSPE coincides with the MSE [40]. It is observed in both figures that, at high SNRs, the performance of the CS-AML coincides with the

[4]The work [39] studies blind estimation of an LPTV CIR applicable to the model of [10, Sec. VI].

[5]Recalling that the PS is non-random and Eqn. (4) we obtain that $\left|y[n] - w[n]\right| = \left|e^{j2\pi\alpha\lfloor\frac{n}{N_0}\rfloor} \sum_{l=0}^{L_{\text{ch}}} \frac{s[n-l]}{\sqrt{N_0}} \sum_{k=0}^{N_0-1} e^{j\frac{2\pi}{N_0}nk} \underset{\sim}{h}[k,l]\right|$, thus, $\|\mathbf{y} - \mathbf{w}\|_2^2$ is deterministic and does not depend on α.

[6]Note that $\mathcal{K}$ depends on the random observations $\mathbf{y}$, hence, $\sum_{k\in\mathcal{K}} |[\underset{\sim}{\mathbf{h}}]_k|^2$ is a random variable.

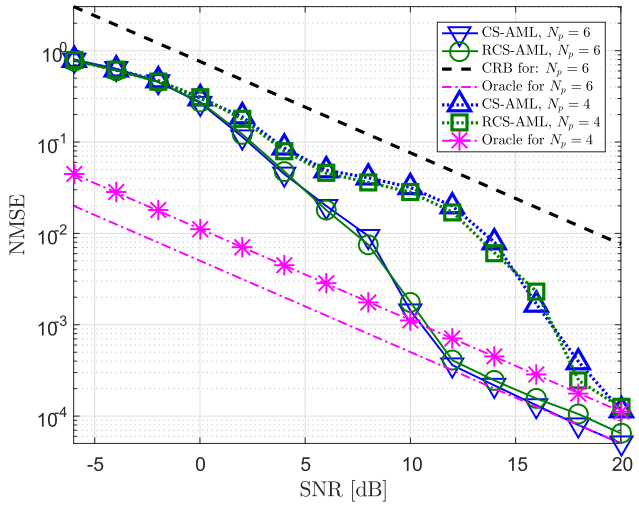


Fig. 2. CIR NMSE, sparse mobile radio scenario.

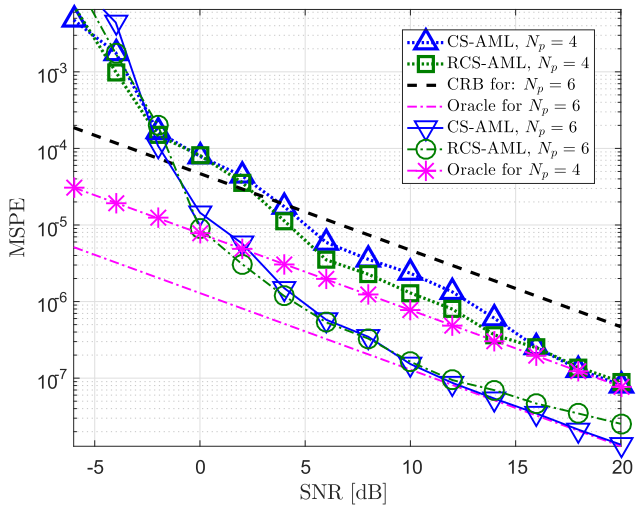


Fig. 3. CFO MSPE, sparse mobile radio scenario.

oracle CRB, as the sparsity pattern estimation successfully identifies the exact sparsity pattern. This demonstrates that when the channel is exactly sparse, the number of non-zero DDSF coefficients is sufficiently small, and the SNR is sufficiently high, then, the CS-AML is capable of achieving the oracle CRB, which lower bounds the estimation performance when the sparsity pattern is known. We also note that the performance of the RCS-AML at high SNRs evaluated for $N_{\mathrm{p}} = 6$ arrives very close to the oracle CRB. The reason for the small gap observed is that the RCS-AML estimator neglects the correlation between observations that belong to subsequent periods. Note that for $N_{\mathrm{p}} = 4$, the RCS-AML seems to coincide with the oracle CRB. This can be explained by noting that for $N_{\mathrm{p}} = 4$, the excess NMSE which results from neglecting the correlation between observed sequences is negligible compared to the overall NMSE, hence it is not visually observed in the figure.

Next, we compare the computational complexity of the tested estimators. Note that the JMLE estimates a total of $N_{\mathrm{CIR}} = N_0 \cdot L_{\mathrm{ch}}$ DDSF coefficients, hence, the computational

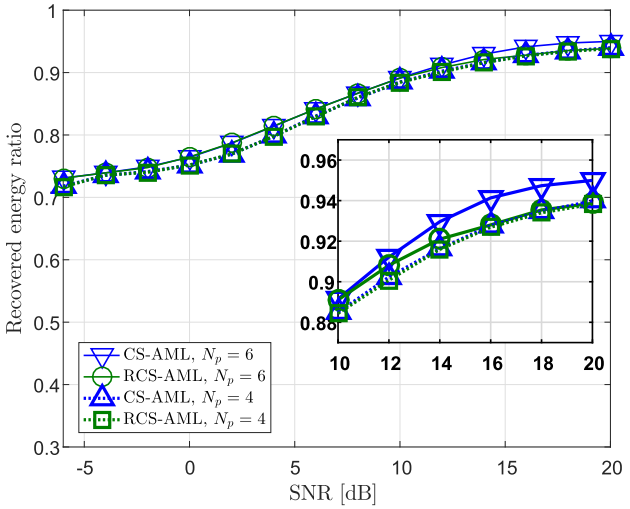


Fig. 4. Performance of sparsity pattern estimation, NB-PLC scenario.

complexity of the JMLE is on the order of $\mathcal{O}\left(T \cdot N_{\mathrm{CIR}}^3\right)$ multiplications and summations. Specifically, for $N_{\mathrm{CIR}} = 100$, the computational complexity of the JMLE is $\mathcal{O}\left(10^{10}\right)$ multiplications and summations. To evaluate the computational complexity of the CS-AML and of the RCS-AML, our numerical study shows that the sparsity pattern estimation algorithm results in average values of $|\mathcal{K}|$ which range from $|\mathcal{K}| = 12$ for $N_{\mathrm{p}} = 4$, and $|\mathcal{K}| = 15$ for $N_{\mathrm{p}} = 6$ at SNR $= 0$ dB, down to the true value of $|\mathcal{K}| = 9$ at SNRs above 10 dB for $N_{\mathrm{p}} = 6$, and at SNRs above 16 dB for $N_{\mathrm{p}} = 4$. Consequently, the computational complexity of the CS-AML at high SNR is on the order of $\mathcal{O}\left(T \cdot |\mathcal{K}|^3\right) \cong \mathcal{O}\left(10^7\right)$ multiplications and summations, while the computational complexity of the RCS-AML is on the order of $\mathcal{O}\left(T \cdot |\mathcal{K}|^2\right) \cong \mathcal{O}\left(10^6\right)$ multiplications and summations. We conclude that at high SNRs, using the CS-AML and the RCS-AML reduces the computational complexity for this scenario by factors of $\sim 10^3$ and $\sim 10^4$ compared to the standard JMLE, respectively.

To summarize, Figs. 2–3 indicate that the proposed algorithms achieve near optimal performances at a considerably lower computational complexity compared to the JMLE. It is therefore interesting to characterize the performance of the proposed algorithms in practical approximately sparse channels, which is done next.

B. Performance Evaluation in the Approximately Sparse Scenarios

We begin with evaluating the performance of the sparsity pattern estimation. The results are depicted for the NB-PLC scenario in Fig. 4, and for the mobile radio scenario in Fig. 5. As observed in Figs. 4–5, the sparsity pattern estimation algorithm successfully identifies the indexes of the dominant CIR coefficients for both $N_{\mathrm{p}} = 4$ and for $N_{\mathrm{p}} = 6$, particularly above moderately high SNRs. For the NB-PLC channel, when $N_{\mathrm{p}} = 6$ we observe from Fig. 4 that the mean ratio values of the RCS-AML are slightly smaller compared to the mean ratios achieved by the CS-AML at high SNRs, which is due to the restriction $|\mathcal{K}| \leq N_0$. Inspecting the sparsity estimation

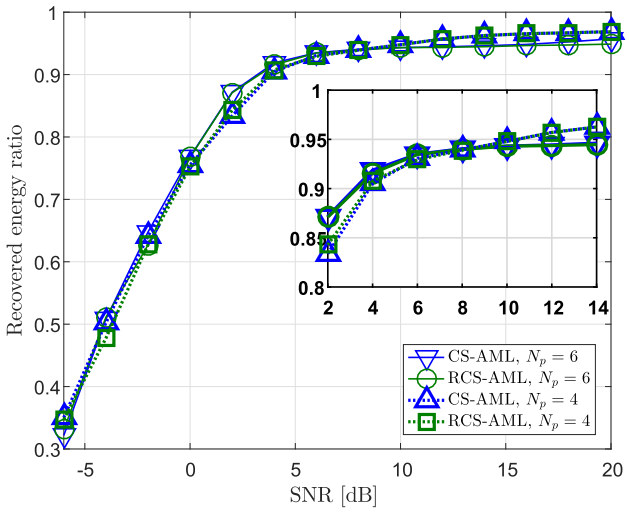


Fig. 5. Performance of sparsity pattern estimation, mobile radio scenario.

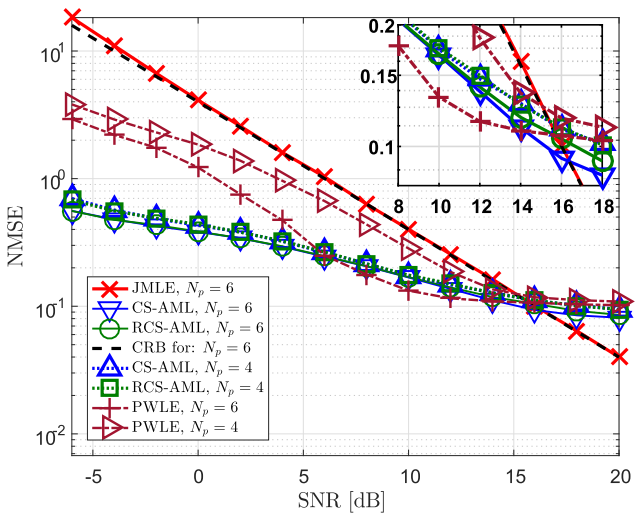


Fig. 6. CIR NMSE, NB-PLC scenario.

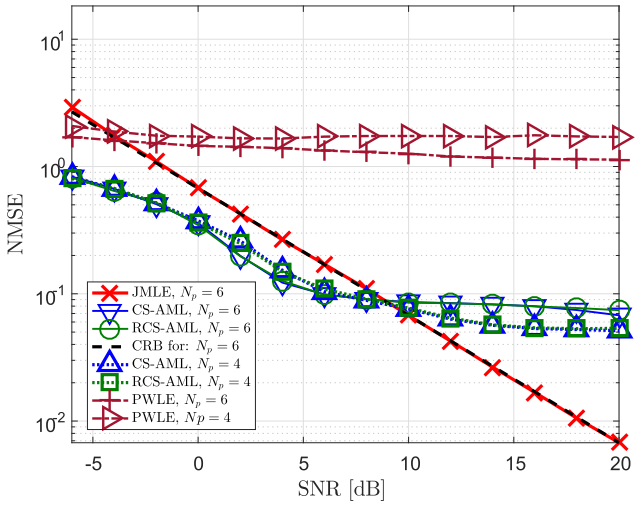


Fig. 7. CIR NMSE, mobile radio scenario.

performance for the mobile radio channel in Fig. 5, we note that at high SNRs the performance for $N_{\mathrm{p}} = 4$ is better than the performance for $N_{\mathrm{p}} = 6$. This situation can be attributed to the lower MC of the PS used for $N_{\mathrm{p}} = 4$ compared to the MC of the PS used for $N_{\mathrm{p}} = 6$. Lastly, we note that for both scenarios, the maximal value of $r(\mathcal{K})$ is approximately $\sim 95\%$, as the DDSF coefficients are only *approximately sparse* (cf. Fig. 1 for the exactly sparse case). Since a certain part of the DDSF coefficients is excluded from the set $\mathcal{K}$ by the sparsity estimation algorithm, it is expected that the NMSE for the estimated CIR coefficients will floor at high SNR values.

Next, we compare the estimation performance for the unknown parameters $\underset{\sim}{\mathbf{h}}$ and α for the CS-AML, RCS-AML, and the JMLE. We also include the CRB when $N_{\mathrm{p}} = 6$ for a reference (when $N_{\mathrm{p}} = 4$ the CRB does not exist since $N_{\mathrm{p}} < L_{\mathrm{ch}}$). In addition, we present the estimation performance for an algorithm obtained by adapting one of the major algorithms for the joint estimation of CFO and CIR for *LTI channels*, proposed in [13], to estimation over linear periodic channels. The algorithm of [13] is adapted to periodic channels by dividing the channel period into N_{I} intervals, and treating the CIR as approximately LTI during each interval. A similar approach was applied in the study of bit loading for periodic channels in [41]. We henceforth refer to the estimator obtained by adapting [13] as the piecewise LTI estimator (PWLE). Specifically, in both scenarios considered, the period is divided into $N_{\mathrm{I}} = 4$ intervals, each consisting of 5 samples. The estimates of the approximately LTI CIR coefficients and of the CFO ν (measured between subsequent samples) are obtained using the set of observations for each interval via the JMLE of [13]. The final estimate of the CFO ν is obtained by averaging over the N_{I} estimates of ν obtained from each interval. Lastly, values of the CIR $g[n,l]$ are obtained $\forall n \in \mathcal{N}_0$ by repeating the estimated L_{ch} LTI coefficients $\frac{N_0}{N_{\mathrm{I}}} = 5$ times (i.e., over a single interval) and then switching to a new LTI CIR with each new interval. Clearly, the performance of the PWLE is limited due to the mismatch between the true channel model and the piecewise LTI model assumed by the PWLE.

The NMSE for the estimation of $\underset{\sim}{\mathbf{h}}$ is[7] depicted in Figs. 6 and 7, for the NB-PLC scenario and the mobile radio scenario, respectively. As observed in Figs. 6–7, while the CRB lower bounds the performance of the *unbiased* JMLE, and is achieved by the JMLE over a wide range of SNR values, the CRB does not lower bound the performance of the CS-AML and the RCS-AML for two main reasons:

1) The CIR estimate is biased since non-zero coefficients excluded from the set $\mathcal{K}$ are estimated as 0. This bias results in an NMSE of $1 - r(\mathcal{K})$. Additionally, the estimated values of coefficients belonging to $\mathcal{K}$ are such that the contribution of non-zero coefficients excluded from $\mathcal{K}$ to the received signal is minimized as detailed in Lemma 3. This effect further increases the bias of the estimates.

[7] As the PWLE estimates $g[n,l]$ instead of $h[n,l]$, its CIR NMSE is given by $\mathbb{E}\left\{\sum_{\substack{n\in\mathcal{N}_0\\ l\in\mathcal{L}_{\mathrm{ch}}}}\left|\hat{g}[n,l]e^{j2\pi\hat{\nu}n} - h[n,l]\right|^2\right\}/\|\underset{\sim}{\mathbf{h}}\|^2$.

2) The estimates of the DDSF coefficients are coupled, since the associated Fisher information matrix (FIM), derived in Appendix D, is in general non-diagonal. Consequently, since the application of the sparse approximation decreases the number of estimated CIR coefficients and introduces bias, the resulting estimated DDSF coefficients are no longer unbiased, which may result in the NMSE for the remaining coefficients becoming smaller than the CRB obtained for estimating the entire DDSF.

As can be observed from Figs. 6–7, the CS-AML and the RCS-AML outperform the standard JMLE over a wide range of SNRs, due to decreasing the number of estimated parameters via sparsity. Specifically, we observe that even when the number of observed periods is decreased, from $N_{\mathrm{p}} = 6$ to $N_{\mathrm{p}} = 4$, the CS-AML and the RCS-AML still outperform the JMLE (for $N_{\mathrm{p}} = 6$) at low and medium SNRs. We therefore conclude that *the proposed estimators can provide a substantial performance gain, in addition to the increased spectral efficiency and reduced computational complexity*. Figs. 6 and 7 also include the performance of the PWLE for comparison with the proposed algorithms. In the NB-PLC scenario depicted in Fig. 6, it is observed that both the CS-AML and the RCS-AML outperform the PWLE for most of the considered SNR range for $N_{\mathrm{p}} = 6$ and for all of the SNR range for $N_{\mathrm{p}} = 4$. This shows that the CS-based estimators can provide better CIR estimation performance at high spectral efficiency compared to the ad-hoc solution. In the mobile radio channel, it can be observed from Fig. 7 that the PWLE is not able to effectively estimate the CIR coefficients. This is because the mobile radio scenario CIR varies rapidly w.r.t. n, hence it cannot be approximated as piecewise LTI with sufficient accuracy. We therefore conclude that *the CS-AML and RCS-AML are applicable to a wider range of communications scenarios than the PWLE*. In addition, from Fig. 7 we observe that for the mobile radio scenario, the CS-based estimators obtain better CIR estimation performance with $N_{\mathrm{p}} = 4$ than with $N_{\mathrm{p}} = 6$ at high SNRs. This follows since higher $r(\mathcal{K})$ values are obtained with $N_{\mathrm{p}} = 4$ than for $N_{\mathrm{p}} = 6$ for this scenario, as was also noted in the context of Fig. 5. This further emphasizes the importance of designing the PS in order to improve the performance of sparsity pattern estimation. Finally, in order to present a complete treatment of the performance of the CS-AML and the RCS-AML estimators for an approximately sparse DDSF, we depict in Fig. 8 the NMSE when the square error and normalization are only w.r.t. the actual DDSF coefficients in $\mathcal{K}$, i.e., $\mathbb{E}\left\{\|\hat{\underset{\sim}{\mathbf{h}}}_{\mathrm{r}} - \underset{\sim}{\mathbf{h}}_{\mathrm{r}}\|^2\right\}/\mathbb{E}\left\{\|\underset{\sim}{\mathbf{h}}_{\mathrm{r}}\|^2\right\}$. As can observed in Fig. 8, when $N_{\mathrm{p}} = 6$, then at high SNRs the NMSE for the estimation of $\underset{\sim}{\mathbf{h}}_{\mathrm{r}}$ via the RCS-AML is lower than via the CS-AML. This observation coincides with the observation that at high SNRs, the RCS-AML estimates a smaller number of DDSF coefficients, noted in the context of Fig. 4, hence *decreasing* $|\mathcal{K}|$ *results in improved estimation accuracy of the coefficients contained in* $\mathcal{K}$. However, as observed from Fig. 6, when comparing the NMSE computed using the square error and normalization w.r.t. the entire CIR vector, then the CS-AML has a lower NMSE than that of the RCS-AML. This is because the CS-AML estimates a larger number of DDSF coefficients compared to the RCS-AML, which in turn increases $r(\mathcal{K})$ and decreases the bias induced by excluding non-zero coefficients from $\mathcal{K}$ for the CS-AML compared to the RCS-AML. Therefore, we conclude that *when the DDSF is only approximately sparse there is a tradeoff between the size of* $|\mathcal{K}|$ *and the estimation performance for the CIR*.

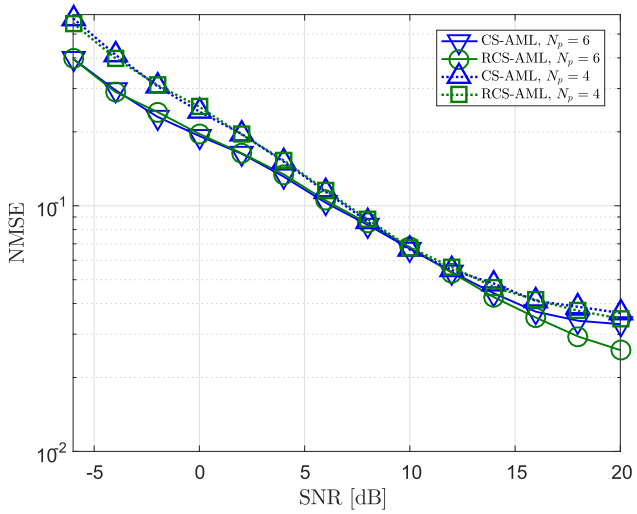


Fig. 8. NMSE of CIR entries in $\mathcal{K}$, NB-PLC scenario.

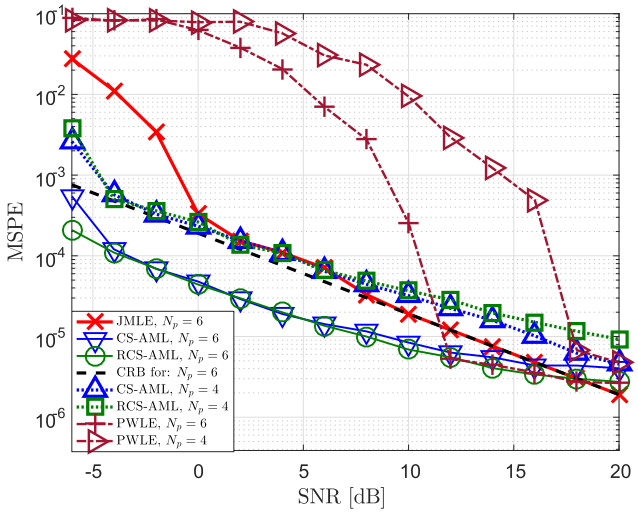


Fig. 9. CFO MSPE, NB-PLC scenario.

The CFO estimation performance is[8] depicted in Figs. 9 and 10 for the NB-PLC scenario and for the mobile radio scenario, respectively. It is observed in Figs. 9–10 that for $N_{\mathrm{p}} = 6$, both the CS-AML and the RCS-AML outperform the JMLE over a wide SNR range. This can be explained by noting that the CFO estimation in (7a), (13a), and (18a), yields the values of α which maximizes the energy of a projection of the measurements $\mathbf{y}$ onto a signal subspace which depends on the tested value of α. For the CS-AML and the RCS-AML, the dimensions of the

[8] Since the PWLE estimates ν instead of α, we define its CFO MSPE as $\mathbb{E}\left\{\left\langle\langle\hat{\nu} N_0\rangle - \alpha\right\rangle^2\right\}$.

TABLE III
COMPUTATIONAL COMPLEXITY OF THE JMLE, CS-AML, RCS-AML, AND PWLE FOR BOTH THE NB-PLC AND MOBILE RADIO SCENARIO

Scenario	SNR	N_{p}	Estimated $\mathbb{E}\{\lvert\mathcal{K}\rvert\}$		Complexity			
			CS-AML	RCS-AML	JMLE	CS-AML	RCS-AML	PWLE
NB-PLC	−6 dB	6	5	5	$\mathcal{O}\left(10^{10}\right)$	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{5}\right)$	$\mathcal{O}\left(10^{6}\right)$
		4	5	5	n/a	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{5}\right)$	$\mathcal{O}\left(10^{6}\right)$
	20 dB	6	28	20	$\mathcal{O}\left(10^{10}\right)$	$\mathcal{O}\left(10^{8}\right)$	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{6}\right)$
		4	20	20	n/a	$\mathcal{O}\left(10^{8}\right)$	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{6}\right)$
Mobile radio	−6 dB	6	9	9	$\mathcal{O}\left(10^{10}\right)$	$\mathcal{O}\left(10^{7}\right)$	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{6}\right)$
		4	9	9	n/a	$\mathcal{O}\left(10^{7}\right)$	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{6}\right)$
	20 dB	6	25	20	$\mathcal{O}\left(10^{10}\right)$	$\mathcal{O}\left(10^{8}\right)$	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{6}\right)$
		4	19	19	n/a	$\mathcal{O}\left(10^{8}\right)$	$\mathcal{O}\left(10^{6}\right)$	$\mathcal{O}\left(10^{6}\right)$

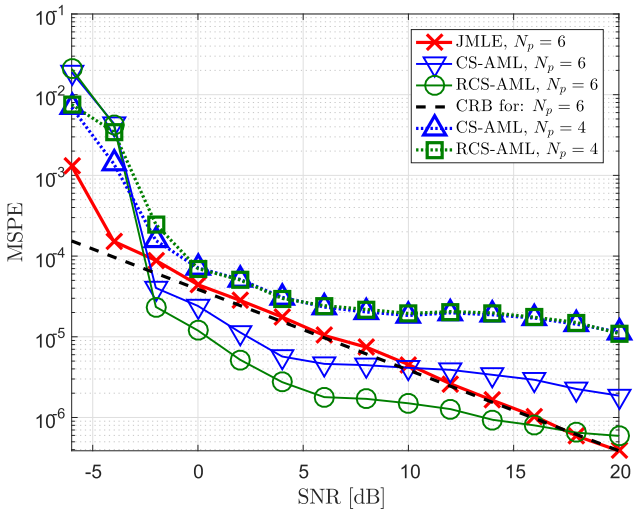


Fig. 10. CFO MSPE, mobile radio scenario.

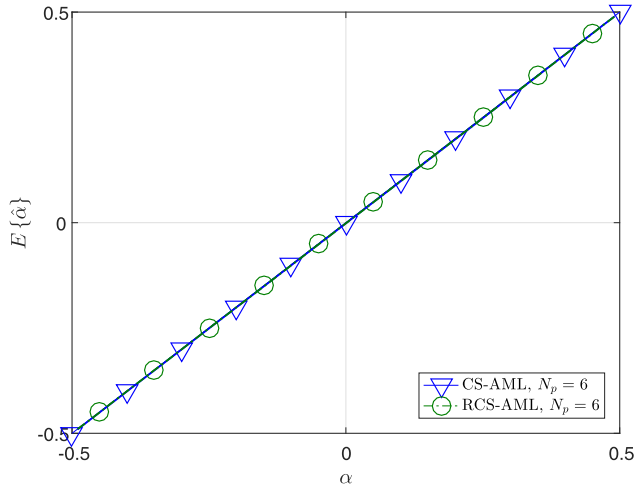


Fig. 11. CFO estimation empirical mean, NB-PLC scenario.

signal subspace is reduced compared the JMLE, as these estimators account for the approximately sparse nature of the DDSF. Consequently, in the CS-AML and the RCS-AML, only signal space dimensions which correspond to dominant DDSF coefficients are used in the estimation of α, resulting in improved performance. Moreover, in some scenarios, the fact that the RCS-AML restricts the number of estimated DDSF coefficients compared to the CS-AML, results in better CFO estimation performance compared to the CS-AML, as can be clearly observed in Fig. 10. Lastly, we observe in Fig. 9 that for the NB-PLC scenario, the CS-AML and RCS-AML considerably outperform the PWLE for most SNR values. For the mobile radio scenario, the CFO MSPE of the PWLE is so much worse than for all other estimators that its MSPE plots are out of scale, hence, they are not included in Fig. 10.

A major property that has to be characterized is the bias of the CFO estimates. To that aim, we depict the mean CFO estimate in Fig. 11 for the NB-PLC scenario with $N_{\mathrm{p}} = 6$, for values of α over the interval $I_{1/2}$ at a fixed SNR of 5 dB. It is illustrated in Fig. 11 that both the CS-AML and the RCS-AML are unbiased for any CFO value in the interval, while for $\alpha = 0.2$, it is observed in Fig. 9 that these estimators achieve an MSPE *which at some SNRs is 8 dB lower than the CRB*. The fact that the CFO estimator is an unbiased estimator whose MSPE is lower than the CRB can be explained by first noting from the FIM (detailed in Appendix D) that the estimation error for unbiased CFO estimation *is coupled* with the estimation error for unbiased estimation of the CIR coefficients. As the application of sparsity pattern estimation results in *biased estimation of the CIR*, the above mentioned coupling allows achieving unbiased CFO estimation whose MSPE is lower than the CRB for CFO estimation.

Lastly, in Table III we compare the computational complexity of the tested estimators for both the NB-PLC scenario and the mobile radio scenario. The table details the empirical mean value of $|\mathcal{K}|$, $\mathbb{E}\{|\mathcal{K}|\}$, for SNR $= -6$ dB and for SNR $= 20$ dB, along with the corresponding computational complexity orders for each estimator. We note that when $|\mathcal{K}|$ is larger than the period N_0, then the RCS-AML estimates only the most dominant N_0 DDSF coefficients. Recall that $N_{\mathrm{CIR}} = N_0 \cdot L_{\mathrm{ch}} = 100$ in both scenarios, hence, the computational complexity of the JMLE is of the order of $\mathcal{O}\left(T \cdot N_{\mathrm{CIR}}^3\right) \cong \mathcal{O}\left(10^{10}\right)$ multiplications and summations for both the NB-PLC scenario and the mobile radio scenario, for all SNR values. The

PWLE computes an estimate of the frequency offset parameter ν for *each* of the $N_{\mathrm{I}} = 4$ intervals. The estimate at each interval is obtained via a grid search over T test values; As each test value requires the inversion of an $L_{\mathrm{ch}} \times L_{\mathrm{ch}}$ matrix (recall that the PWLE estimates $L_{\mathrm{ch}} = 5$ CIR coefficients for each interval) it follows that the computational complexity of the PWLE is on the order of $\mathcal{O}\left(N_{\mathrm{I}} \cdot T \cdot L_{\mathrm{ch}}^3\right) \cong \mathcal{O}\left(10^6\right)$ multiplications and summations.

As evident from Table III, both the CS-AML and the RCS-AML have a *considerably smaller* computational complexity than that of the JMLE for all SNR values considered. Since $\mathbb{E}\{|\mathcal{K}|\}$ decreases as the SNR decreases, it follows that as the SNR decreases, then, the computational complexity of the CS-AML and of the RCS-AML decreases compared to the computational complexity of the JMLE, which remains constant. Note that the fact that the number of estimated DDSF coefficients increases as the SNR increases is expected, since the parameter σ^2, used by the sparsity estimation algorithm, depends linearly on the expected power of the noise. As the SNR increases, the parameter σ^2 decreases, and consequently, the constraint imposed on the sparsity pattern estimation algorithm. i.e., $||\mathbf{y} - \mathbf{S}\mathbf{F}\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)||^2 \leq \sigma^2$, leads to determining an increasing number of groups of coefficients from $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ to be non-zero, which in turn results in an increase in $|\mathcal{K}|$. Note that while in the approximately sparse scenario, $\mathbb{E}\{|\mathcal{K}|\}$ increases as the SNR increases, for the exactly sparse scenario, $\mathbb{E}\{|\mathcal{K}|\}$ decreases as the SNR increases. The reason is that in the exactly sparse scenario and at high SNR, the constraint $||\mathbf{y} - \mathbf{S}\mathbf{F}\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)||^2 \leq \sigma^2$ can be satisfied with high probability by determining coefficients of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ as non-zero only from the groups associated with the actual non-zero coefficients of $\underset{\sim}{\mathbf{h}}$. Thus, when the SNR is high, the BPDN algorithm correctly determines the 9 non-zero groups of coefficients of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$, which results in $\mathbb{E}\{|\mathcal{K}|\} = 9$. Conversely, at low and at intermediate SNRs, when the noise is more dominant, the BPDN algorithm declares also groups of zero coefficients of $\underset{\sim}{\mathbf{h}}_{\mathbf{\Phi}}(\alpha)$ to be non-zero, which results in an increase in the number of estimated coefficients.

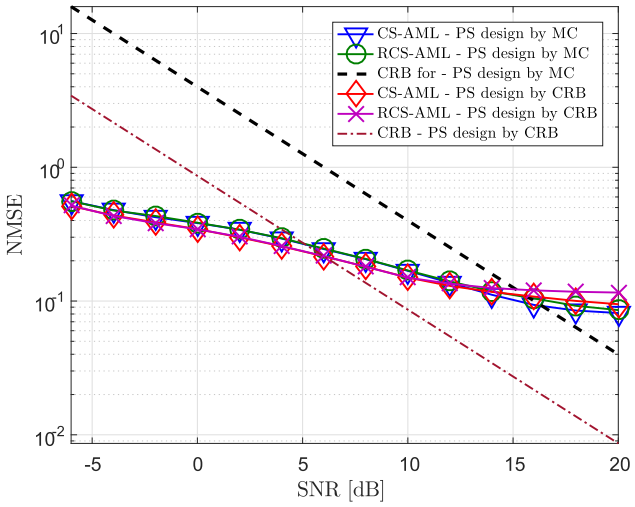


Fig. 12. NMSE for CIR estimation for PS design by MC minimization and for PS design by CRB minimization, NB-PLC scenario.

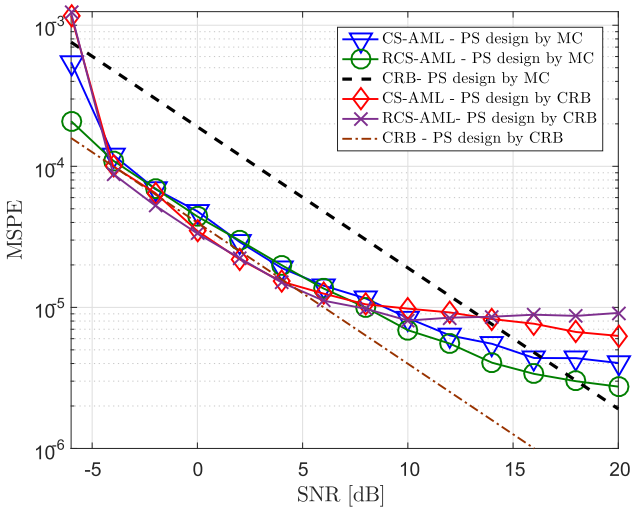


Fig. 13. MSPE for CFO estimation for PS design by MC minimization and for PS design by CRB minimization, NB-PLC scenario.

C. Performance Evaluation of the Proposed PS Design Approach

Lastly, we numerically evaluate the proposed PS design approach based on minimizing the MC, compared to PS design based on minimizing the trace of the CRB for the estimation of the CIR coefficients, proposed in [17] and [20], *which requires a-priori knowledge of the unknown parameters* $\underset{\sim}{\mathbf{h}}$ *and* α. Both PSs are designed subject to the constraints stated in (23b). For the PS design scheme which minimizes the trace of the CRB, we used the true values of the CIR and CFO which are, in fact, unknown in our scenario. The performance of the CS-AML and of the RCS-AML for both PSs for the NB-PLC scenario is depicted in Figs. 12-13: Fig. 12 depicts the NMSE for CIR estimation and Fig. 13 depicts the MSPE for CFO estimation. It is observed in Fig. 12 that designing the PS to minimize the CRB for the estimation of the CIR yields a substantial improvement in the CRB, compared to PS designed to minimize the MC. A similar improvement is also noted in the CRB for the CFO estimation, due to the strong dependence between the CRBs, see (33) in Appendix D. Nonetheless, it is demonstrated in Figs. 12-13 that the CRB improvement does not translate into a performance gain for the CS-AML and the RCS-AML. This follows since sparsity pattern estimation plays a crucial role in the CS-AML and the RCS-AML estimators, hence the estimators with the PS design based on MC minimization, which facilitates the sparsity pattern estimation, obtain roughly the same performance as with the PS design based on CRB minimization, *without requiring a-priori knowledge of the unknown parameters*. These results illustrate the gain of the proposed PS design guidelines.

VI. Conclusions

In this paper, a novel joint CIR and CFO estimation scheme for channels with periodic characteristics is proposed. The estimation scheme offers improved spectral efficiency by

exploiting the sparse nature of the DDSF coefficients of the CIR using CS-based techniques, and a simplified version of the scheme can be implemented with a reduced computational complexity at the cost of a minor performance degradation. Comprehensive complexity analysis of the proposed schemes is provided, and guidelines for PS design are presented. The performance of the proposed schemes is evaluated in a simulation study corresponding to practical scenarios, and is compared with that of the JMLE for each scenario and with that of a reference ad-hoc estimation scheme obtained based on adapting an estimation scheme for LTI channels. The numerical results demonstrate that the proposed estimation schemes are applicable to various communications scenarios, and illustrate that properly exploiting the sparsity of the channel facilitates improved estimation performance, particularly at low and medium SNRs.

APPENDIX

A. Proof of Lemma 1

Substituting the definition of $h[n,l]$ into (3) we obtain

$$\underset{\sim}{h}[k,l]=\frac{1}{\sqrt{N_0}}\sum_{n=0}^{N_0-1}g[n,l]e^{j\phi_0}e^{j\frac{2\pi}{N_0}((n))_{N_0}N_0\nu}e^{-j\frac{2\pi}{N_0}nk}, \quad (24)$$

i.e., $\forall l \in \mathcal{L}_{\rm ch}$, $\underset{\sim}{h}[k,l]$ is the DFS with respect to n of the product between $g[n,l]$ and the periodic function $v[n] \triangleq e^{j\phi_0}e^{j\frac{2\pi}{N_0}N_0\nu((n))_{N_0}}$. Applying the properties of the DFS [31, Table 8.2] we obtain that $\underset{\sim}{h}[k,l]$ can be obtained as the circular convolution between $\underset{\sim}{g}[k,l]$ and the DFS of $v[n]$. Let $V[k]$ denote the DFS of $v[n]$. Then, directly applying the DFS formula to $v[n]$, we obtain an explicit expression for $V[k]$:

$$V[k]=\frac{e^{j\phi_0}}{\sqrt{N_0}}\frac{\sin\left(\pi(k-N_0\nu)\right)}{\sin\left(\frac{\pi(k-N_0\nu)}{N_0}\right)}e^{-j\pi\frac{(N_0-1)}{N_0}(k-N_0\nu)}, \ k\in\mathcal{N}_0.$$

It therefore follows that $\underset{\sim}{h}[k,l]$ is given by

$$\underset{\sim}{h}[k,l]=\frac{e^{j\phi_0}}{N_0}\sum_{k'=0}^{N_0-1}\underset{\sim}{g}[k',l]\frac{\sin\left(\pi(k-k'-N_0\nu)\right)}{\sin\left(\frac{\pi(k-k'-N_0\nu)}{N_0}\right)}\cdot \cdot e^{-j\pi\frac{(N_0-1)}{N_0}(k-k'-N_0\nu)}, \quad k\in\mathcal{N}_0. \quad (25)$$

Equaling (25) to zero we obtain Condition 2. Note that when $N_0\nu\in\mathbb{Z}$, then the sum in (25) represents the DFS of $g[n,l]$, circularly shifted by $N_0\nu$ frequency bins [31, Ch. 8.2.2], i.e., $\underset{\sim}{h}[k,l]=e^{j\phi_0}\underset{\sim}{g}\left[((k-N_0\nu))_{N_0},l\right]$, from which we obtain Condition 1. This proves the lemma. □

B. Proof of Lemma 3

The KLD between $f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})$ and $f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\rm r})$ is given by

$$\begin{aligned}
&f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})\big|\big|f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\rm r})\\
&=\mathbb{E}_{f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})}\left\{\log\frac{f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})}{f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\rm r})}\right\}\\
&\overset{(a)}{=}\mathbb{E}_{f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})}\Big\{\underset{\sim}{\mathbf{h}}_{\rm r}^H\mathbf{H}_{{\rm r},(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}-\underset{\sim}{\mathbf{h}}^H\mathbf{H}_{(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}\\
&\quad+\mathbf{y}^H\mathbf{C}_{\mathbf{w}}^{-1}\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}-\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}\right)+\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}-\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}\right)^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}\Big\}\\
&\overset{(b)}{=}\underset{\sim}{\mathbf{h}}^H\mathbf{H}_{(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}-\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}\right)\\
&\quad+\underset{\sim}{\mathbf{h}}_{\rm r}^H\mathbf{H}_{{\rm r},(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\left(\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}-\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}\right), \qquad (26)
\end{aligned}$$

where (a) follows since both distributions are Gaussian and have the same covariance matrix and (b) follows since $\mathbb{E}_{f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})}\{\mathbf{y}\}=\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}$. Differentiating (26) w.r.t. $\underset{\sim}{\mathbf{h}}_{\rm r}^H$ (see [32, Sec. 15.6]), and letting $\underset{\sim}{\mathbf{h}}_{\rm r}^{\rm opt}$ denote the value of $\underset{\sim}{\mathbf{h}}_{\rm r}$ for which the derivative equals $\mathbf{0}_{|\mathcal{K}|\times 1}$, we obtain $\underset{\sim}{\mathbf{h}}_{\rm r}^{\rm opt}=\mathbf{G}_{{\rm r},(\alpha)}\mathbf{H}_{{\rm r},(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}$. Note that the Hessian of (26) is the positive definite matrix $\mathbf{H}_{{\rm r},(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}_{{\rm r},(\alpha)}$, hence, $\underset{\sim}{\mathbf{h}}_{\rm r}^{\rm opt}$ is a global minima. Substituting $\underset{\sim}{\mathbf{h}}_{\rm r}^{\rm opt}$ into (26) we obtain that the minimal value of the KLD is given by:

$$\min_{\underset{\sim}{\mathbf{h}}_{\rm r}\in\mathbb{C}^{|\mathcal{K}|}}D\Big(f_0(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}})\big|\big|f_1(\mathbf{y};\alpha,\underset{\sim}{\mathbf{h}}_{\rm r})\Big)=\left\|\mathbf{B}_{{\rm r},(\alpha)}^{\perp}\mathbf{C}_{\mathbf{w}}^{-1/2}\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}\right\|_2^2,$$

which equals (14b). Lastly, by substituting $\hat{\alpha}^{\rm CS}=\alpha$ in (13b), carrying out the stochastic expectation and writing $\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}=\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}-\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}+\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}$, we indeed obtain:

$$\begin{aligned}
\mathbb{E}\left\{\underset{\sim}{\hat{\mathbf{h}}}_{{\rm r},(\alpha)}^{\rm CS}\right\}&=\mathbf{G}_{{\rm r},(\alpha)}\mathbf{H}_{{\rm r},(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}=\underset{\sim}{\mathbf{h}}_{\rm r}^{\rm opt}\\
&=\underset{\sim}{\mathbf{h}}_{\rm r}+\mathbf{G}_{{\rm r},(\alpha)}\mathbf{H}_{{\rm r},(\alpha)}^H\mathbf{C}_{\mathbf{w}}^{-1}\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}-\mathbf{H}_{{\rm r},(\alpha)}\underset{\sim}{\mathbf{h}}_{\rm r}\right), \qquad (27)
\end{aligned}$$

which completes the proof of Lemma 3. □

C. Derivation of Eqns. (18)

The gradient of (17) w.r.t. $\underset{\sim}{\mathbf{h}}_{\rm r}$ is given by its derivative w.r.t. $\underset{\sim}{\mathbf{h}}_{\rm r}^H$ [32, Sec. 15.6]. Carrying out the differentiation and setting the derivative to zero, we obtain:

$$\begin{aligned}
&\sum_{p=0}^{N_{\rm p}-1}e^{-j2\pi\alpha p}\underset{\sim}{\mathbf{C}}_{{\rm r},p}^{-1}\left(e^{j2\pi\alpha p}\underset{\sim}{\mathbf{h}}_{\rm r}-\underset{\sim}{\hat{\mathbf{h}}}_{{\rm r},p}^{(\alpha)}\right)\\
&\qquad=\sum_{p=0}^{N_{\rm p}-1}\underset{\sim}{\mathbf{C}}_{{\rm r},p}^{-1}\underset{\sim}{\mathbf{h}}_{\rm r}-\sum_{p=0}^{N_{\rm p}-1}\underset{\sim}{\mathbf{C}}_{{\rm r},p}^{-1}\underset{\sim}{\hat{\mathbf{h}}}_{{\rm r},p}^{(\alpha)}e^{-j2\pi\alpha p}=0.
\end{aligned}$$

Since $\underset{\sim}{\mathbf{h}}_{\rm r}$ does not depend on p, it follows that

$$\begin{aligned}
\underset{\sim}{\hat{\mathbf{h}}}_{{\rm r},(\alpha)}^{\rm RCS}&=\sum_{p=0}^{N_{\rm p}-1}\left(\sum_{p'=0}^{N_{\rm p}-1}\underset{\sim}{\mathbf{C}}_{{\rm r},p'}^{-1}\right)^{-1}\underset{\sim}{\mathbf{C}}_{{\rm r},p}^{-1}\underset{\sim}{\hat{\mathbf{h}}}_{{\rm r},p}^{(\alpha)}e^{-j2\pi\alpha p}\\
&=\mathbf{C}_{\rm inv}\sum_{p=0}^{N_{\rm p}-1}\underset{\sim}{\mathbf{C}}_{{\rm r},p}^{-1}\underset{\sim}{\hat{\mathbf{h}}}_{{\rm r},p}^{(\alpha)}e^{-j2\pi\alpha p}. \qquad (28)
\end{aligned}$$

Note that when $\hat{\alpha}^{\rm RCS}$ is substituted instead of α, then (28) coincides with (18b). Next, define the function $\mathbf{x}[p] \triangleq \underset{\sim}{\mathbf{C}}_{{\rm r},p}^{-1}\underset{\sim}{\hat{\mathbf{h}}}_{{\rm r},p}^{(\alpha)}$ for $p\in\mathcal{N}_{\rm p}$ and $\mathbf{x}[p]=0$ otherwise. Note that

$$\mathbf{J} = 2\begin{bmatrix} \overline{\mathbf{s}}^H_{(\alpha,\underset{\sim}{\mathbf{h}})}\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})} & \mathrm{Im}\left\{(\overline{\mathbf{s}}^H_{(\alpha,\underset{\sim}{\mathbf{h}})}\mathbf{C}_{\mathbf{w}}^{-1/2}\mathbf{H}_{(\alpha)}\right\} & \mathrm{Re}\left\{\overline{\mathbf{s}}^H_{(\alpha,\underset{\sim}{\mathbf{h}})}\mathbf{C}_{\mathbf{w}}^{-1/2}\mathbf{H}_{(\alpha)}\right\} \\ -\mathrm{Im}\left\{\mathbf{H}^H_{(\alpha)}\mathbf{C}_{\mathbf{w}}^{-1/2}\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})}\right\} & \mathrm{Re}\left\{\mathbf{G}^{-1}_{(\alpha)}\right\} & -\mathrm{Im}\left\{\mathbf{G}^{-1}_{(\alpha)}\right\} \\ \mathrm{Re}\left\{\mathbf{H}^H_{(\alpha)}\mathbf{C}_{\mathbf{w}}^{-1/2}\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})}\right\} & \mathrm{Im}\left\{\mathbf{G}^{-1}_{(\alpha)}\right\} & \mathrm{Re}\left\{\mathbf{G}^{-1}_{(\alpha)}\right\} \end{bmatrix}. \tag{31}$$

$$\mathbb{E}\left\{(\hat{\alpha}-\alpha)^2\right\} \geq \frac{1}{2}\left(\overline{\mathbf{s}}^H_{(\alpha,\underset{\sim}{\mathbf{h}})}\mathbf{B}^{\perp}_{(\alpha)}\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})}\right)^{-1} \triangleq \mathrm{CRB}(\alpha) \tag{32a}$$

$$\mathbb{E}\left\{(\hat{\underset{\sim}{\mathbf{h}}}-\underset{\sim}{\mathbf{h}})(\hat{\underset{\sim}{\mathbf{h}}}-\underset{\sim}{\mathbf{h}})^H\right\} \geq \mathbf{G}_{(\alpha)} + \mathrm{CRB}(\alpha)\mathbf{G}_{(\alpha)}\mathbf{H}^H_{(\alpha)}\mathbf{C}_{\mathbf{w}}^{-1/2}\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})}\overline{\mathbf{s}}^H_{(\alpha,\underset{\sim}{\mathbf{h}})}\mathbf{C}_{\mathbf{w}}^{-1/2}\mathbf{H}_{(\alpha)}\mathbf{G}_{(\alpha)}. \tag{32b}$$

$$\mathbb{E}\left\{\|\hat{\underset{\sim}{\mathbf{h}}}-\underset{\sim}{\mathbf{h}}\|^2\right\} \geq \mathrm{Trace}\Big(\mathbf{G}_{(\alpha)} + \mathrm{CRB}(\alpha)\mathbf{G}_{(\alpha)}\mathbf{H}^H_{(\alpha)}\mathbf{C}_{\mathbf{w}}^{-1/2}\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})}\overline{\mathbf{s}}^H_{(\alpha,\underset{\sim}{\mathbf{h}})}\mathbf{C}_{\mathbf{w}}^{-1/2}\mathbf{H}_{(\alpha)}\mathbf{G}_{(\alpha)}\Big). \tag{33}$$

$\tilde{\mathbf{x}}(\alpha)$ defined in (19b) is the discrete-time Fourier transform of $\{\mathbf{x}[p]\}_{p=-\infty}^{\infty}$. Using these definitions, substituting (28) into (17) and removing any summands which do not depend on α, we obtain

$$\begin{aligned}\hat{\alpha}^{\mathrm{RCS}} = \underset{\alpha\in I_{1/2}}{\mathrm{argmax}} \Bigg\{ & 2\sum_{p_1=0}^{N_{\mathrm{p}}-1}\sum_{p_2=0}^{N_{\mathrm{p}}-1}\Big(\mathbf{x}[p_1]\Big)^H\mathbf{C}_{\mathrm{inv}}\mathbf{x}[p_2]\,e^{j2\pi\alpha\ (p_1-p_2)} \\ & - \sum_{p_1=0}^{N_{\mathrm{p}}-1}\sum_{p_2=0}^{N_{\mathrm{p}}-1}\Big(\mathbf{x}[p_2]\Big)^H e^{j2\pi\alpha p_2}\mathbf{C}_{\mathrm{inv}}\underset{\sim}{\mathbf{C}}^{-1}_{\mathrm{r},p_1}\mathbf{C}_{\mathrm{inv}} \\ & \cdot \sum_{p_3=0}^{N_{\mathrm{p}}-1}\mathbf{x}[p_3]\,e^{-j2\pi\alpha p_3}\Bigg\}. \end{aligned} \tag{29}$$

$$\begin{aligned} &\overset{(a)}{=} \underset{\alpha\in I_{1/2}}{\mathrm{argmax}} \left\{ \sum_{p_1=0}^{N_{\mathrm{p}}-1}\sum_{p_2=0}^{N_{\mathrm{p}}-1}\Big(\mathbf{x}[p_1]\Big)^H\mathbf{C}_{\mathrm{inv}}\mathbf{x}[p_2]\,e^{j2\pi\alpha\ (p_1-p_2)}\right\} \\ &= \underset{\alpha\in I_{1/2}}{\mathrm{argmax}} \left\{ \Big(\tilde{\mathbf{x}}(\alpha)\Big)^H\mathbf{C}_{\mathrm{inv}}\tilde{\mathbf{x}}(\alpha)\right\}, \end{aligned} \tag{30}$$

which coincides with (18a). Note that (a) follows since $\sum_{p_1=0}^{N_{\mathrm{p}}-1}\mathbf{C}^{-1}_{\mathrm{r},p_1} = \mathbf{C}^{-1}_{\mathrm{inv}}$, hence, the summation over p_1 in the last summand of (29) can be replaced with $\mathbf{C}^{-1}_{\mathrm{inv}}$, and the last summand in (29) equals to the first summand divided by 2. □

D. CRB Derivation for Joint Estimation of the CIR Coefficients and the CFO

The CRB derivation follows similar steps to the CRB derivation in [17] and [20], with the exception that [17] and [20] considered AWGN while we consider temporally correlated noise. To facilitate the derivation of the CRB, we define the following notations and quantities:

1) $\underset{\sim}{\mathbf{h}}_{\mathrm{R}} \triangleq \mathrm{Re}\{\underset{\sim}{\mathbf{h}}\}$, $\underset{\sim}{\mathbf{h}}_{\mathrm{I}} \triangleq \mathrm{Im}\{\underset{\sim}{\mathbf{h}}\}$, and the real parameter vector $\mathbf{v} \triangleq \left[\alpha, \underset{\sim}{\mathbf{h}}^T_{\mathrm{R}}, \underset{\sim}{\mathbf{h}}^T_{\mathrm{I}}\right]^T$.
2) The diagonal $\mathbf{P} \in \mathbb{C}^{N_{\mathrm{p}}N_{\mathrm{CIR}}\times N_{\mathrm{p}}N_{\mathrm{CIR}}}$ such that $[\mathbf{P}]_{pN_{\mathrm{CIR}}+k,pN_{\mathrm{CIR}}+k} \triangleq 2\pi p$, $\forall p \in \mathcal{N}_{\mathrm{p}}$, $k \in \mathcal{N}_{\mathrm{CIR}}$.
3) The matrix $\mathbf{H}'_{(\alpha)} \triangleq \mathbf{SFP}\mathbf{\Phi}_{(\alpha)}$. Hence, we write $\frac{d}{d\alpha}\mathbf{H}_{(\alpha)} = \mathbf{SF}\cdot\left(\frac{d}{d\alpha}\mathbf{\Phi}_{(\alpha)}\right) = j\mathbf{SFP}\mathbf{\Phi}_{(\alpha)} = j\mathbf{H}'_{(\alpha)}$.
4) The $N_{\mathrm{obs}} \times N_{\mathrm{obs}}$ projection matrix $\mathbf{B}_{(\alpha)} \triangleq \mathbf{C}_{\mathbf{w}}^{-1/2}\mathbf{H}_{(\alpha)}\mathbf{G}_{(\alpha)}\mathbf{H}^H_{(\alpha)}\mathbf{C}_{\mathbf{w}}^{-1/2}$.
5) The vector $\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})} \in \mathbb{C}^{N_{\mathrm{obs}}}$ such that $\overline{\mathbf{s}}_{(\alpha,\underset{\sim}{\mathbf{h}})} \triangleq \mathbf{C}_{\mathbf{w}}^{-1/2}\mathbf{H}'_{(\alpha)}\underset{\sim}{\mathbf{h}}$. Using these definitions we have $\mathbf{y} = \mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}_{\mathrm{R}} + j\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}_{\mathrm{I}} + \mathbf{w}$, $\mathbf{y} \sim \mathcal{CN}\Big(\mathbf{H}_{(\alpha)}(\underset{\sim}{\mathbf{h}}_{\mathrm{R}} + j\underset{\sim}{\mathbf{h}}_{\mathrm{I}}), \mathbf{C}_{\mathbf{w}}\Big)$, and the derivative of the mean of $\mathbf{y}$ w.r.t. $\mathbf{v}$ is given by the concatenation of $j\mathbf{H}'_{(\alpha)}(\underset{\sim}{\mathbf{h}}_{\mathrm{R}} + j\underset{\sim}{\mathbf{h}}_{\mathrm{I}})$, $\mathbf{H}_{(\alpha)}$ and $j\mathbf{H}_{(\alpha)}$, i.e., $\frac{d\left(\mathbf{H}_{(\alpha)}(\underset{\sim}{\mathbf{h}}_{\mathrm{R}}+j\underset{\sim}{\mathbf{h}}_{\mathrm{I}})\right)}{d\mathbf{v}} = \left[j\mathbf{H}'_{(\alpha)}(\underset{\sim}{\mathbf{h}}_{\mathrm{R}} + j\underset{\sim}{\mathbf{h}}_{\mathrm{I}}), \mathbf{H}_{(\alpha)}, j\mathbf{H}_{(\alpha)}\right]$. Replacing $(\underset{\sim}{\mathbf{h}}_{\mathrm{R}} + j\underset{\sim}{\mathbf{h}}_{\mathrm{I}}) = \underset{\sim}{\mathbf{h}}$, the FIM for the estimation of $\mathbf{v}$ from $\mathbf{y}$, denoted $\mathbf{J}$, is given by $\mathbf{J} = 2\mathrm{Re}\left\{\frac{d\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}\right)^H}{d\mathbf{v}}\mathbf{C}_{\mathbf{w}}^{-1}\frac{d\left(\mathbf{H}_{(\alpha)}\underset{\sim}{\mathbf{h}}\right)}{d\mathbf{v}}\right\}$ [32, Definition 15.52], which is stated explicitly in (31), as shown at the top of this page. From (31), following the same mathematical steps as in [20, Appendix I], we conclude that the CRB for α and $\underset{\sim}{\mathbf{h}}$ is given by (32a) and (32b), as shown at the top of this page, respectively. Lastly, using (32b) we obtain the lower bound on the norm of the coefficients error vector stated in (33), as shown at the top of this page. Note that in the LTI case, i.e., $N_0 = 1$, Eqns. (32b)–(33) specialize the CRB and squared norm of the CIR error vector derived in [20], after properly extending the AWGN result of [20] to the correlated noise case. In particular, we note that differently from the AWGN case considered in [20], when $\mathbf{C}_{\mathbf{w}}^{-1}$ is not diagonal, then, the CRB for both α and $\underset{\sim}{\mathbf{h}}$ depends on α.

REFERENCES

[1] M. Nassar, J. Lin, Y. Mortazavi, A. Dabak, I. H. Kim, and B. L. Evans, "Local utility power line communications in the 3–500 kHz band: Channel impairments, noise, and standards," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 116–127, Aug. 2012.

[2] F. J. C. Corripio, J. A. C. Arrabal, L. D. del Rio, and J. T. E. Munoz, "Analysis of the cyclic short-term variation of indoor power line channels," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 7, pp. 1327–1338, Jul. 2006.

[3] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Process.*, vol. 86, no. 4, pp. 639–697, Apr. 2006.

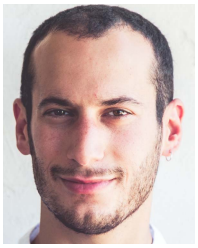
[4] P. Yang, Y. Yan, X.-Y. Li, Y. Zhang, Y. Tao, and L. You, "Taming cross-technology interference for Wi-Fi and ZigBee coexistence networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 4, pp. 1009–1021, Apr. 2016.

[5] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[6] J. Campbell, A. Gibbs, and B. Smith, "The cyclostationary nature of crosstalk interference from digital signals in multipair cable—Part II: Applications and further results," *IEEE Trans. Commun.*, vol. 31, no. 5, pp. 638–649, May 1983.

[7] X. Hong, Z. Chen, C. X. Wang, S. A. Vorobyov, and J. S. Thompson, "Cognitive radio networks," *IEEE Veh. Technol. Mag.*, vol. 4, no. 4, pp. 76–84, Dec. 2009.

[8] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, Oct. 1998.
[9] M. K. Tsatsanis and G. B. Giannakis, "Modelling and equalization of rapidly fading channels," *Int. J. Adapt. Control Signal Process.*, vol. 10, nos. 2–3, pp. 159–176, Mar. 1996.
[10] M. K. Tsatsanis and G. B. Giannakis, "Equalization of rapidly fading channels: Self-recovering methods," *IEEE Trans. Commun.*, vol. 44, no. 5, pp. 619–630, May 1996.
[11] X. Dai, "Optimal training design for linearly time-varying MIMO/OFDM channels modelled by a complex exponential basis expansion," *IET Communi.*, vol. 1, no. 5, pp. 945–953, Oct. 2007.
[12] Y. Jiang, X. Zhu, E. G. Lim, Y. Huang, and H. Lin. *Semi-Blind Carrier Frequency Offset Estimation and Channel Equalization*. Cham, Switzerland: Springer, 2015.
[13] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580–1589, Sep. 2000.
[14] M.-O. Pun, M. Morelli, and C.-C.J. Kuo, "Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2004.
[15] M. Morelli, C. C. J. Kuo, and M. O. Pun, "Synchronization techniques for orthogonal frequency division multiple access (OFDMA): A tutorial review," *Proc. IEEE*, vol. 95, no. 7, pp. 1394–1427, Jul. 2007.
[16] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: A survey," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1673–1694, May 2009.
[17] M. Ghogho and A. Swami, "Training design for multipath channel and frequency-offset estimation in MIMO systems," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.
[18] Y.-H. Chung and S.-M. Phoong, "Joint estimation of I/Q imbalance, CFO and channel response for MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1485–1492, May 2010.
[19] R. Shaked, N. Shlezinger, and R. Dabora, "Carrier frequency offset estimation for linear channels with periodic characteristics," in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Edinburgh, U.K., Jul. 2016, pp. 1–5.
[20] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910–1917, Nov. 2003.
[21] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Urbana-Champaign, IL, USA, Sep. 2008, pp. 575–582.
[22] Y. C. Eldar and G. Kutyniok, Eds., *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[23] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
[24] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
[25] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
[26] G. Tauböck, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, Apr. 2010.
[27] Y. Xu, J. Zhong, Y. Cai, M. Zhao, and B. Chen, "Joint CFO and sparse channel estimation for MIMO-OFDM systems via the SAGE algorithm," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Hangzhou, China, Oct. 2013, pp. 1–5.
[28] R. F. Şenyuva, G. K. Kurt, and E. Anarim, "Compressive sensing based joint frequency offset and channel estimation for OFDM," *EURASIP J. Wireless Commun. Netw.*, vol. 86, Dec. 2016.
[29] F. Gianaroli, F. Pancaldi, and G. M. Vitetta, "On the use of Zadeh's series expansion for modeling and estimation of indoor powerline channels," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2558–2568, Jul. 2014.
[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.
[31] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.
[32] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[33] E. van den Berg, M. Schmidt, M. P. Friedlander, and K. Murphy, "Group sparsity via linear-time projection," Dept. Comput. Sci., Univ. British Columbia, Tech. Rep. TR-2008-09, 2008.
[34] E. van den Berg and M. P. Friedlander, "Sparse optimization with least-squares constraints," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1201–1229, 2011.
[35] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
[36] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.
[37] N. Shlezinger and R. Dabora, "On the capacity of narrowband PLC channels," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1191–1201, Apr. 2015.
[38] *P1901.2/D0.09.00 Draft Standard for Low Frequency (Less Than 500 kHz) Narrow Band Power Line Communications for Smart Grid Applications*, IEEE Standards Association P1901.2/D0.09.00, Jun. 2013.
[39] A. Dandawate and G. B. Giannakis, "Modeling (almost) periodic moving average processes using cyclic statistics," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 673–684, Mar. 1996.
[40] T. Routtenberg and J. Tabrikian, "Non-Bayesian periodic Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 1019–1032, Feb. 2013.
[41] M. A. Tunc, E. Perrins, and L. Lampe, "Optimal LPTV-aware bit loading in broadband PLC," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5152–5162, Dec. 2013.



Roee Shaked (S'16) received the B.Sc. degree in electrical and computer engineering from Ben-Gurion University, Israel, in 2016, where he is currently pursuing the M.Sc. degree in electrical and computer engineering. From 2014 to 2015, he was an Engineer with Yitran Technologies. His research interests include signal processing for communications and compressed sensing.

Nir Shlezinger (S'14) received the B.Sc. and M.Sc. degrees in electrical and computer engineering from Ben-Gurion University, Israel, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering. From 2009 to 2013, he was an Engineer with Yitran Communications. His research interests include information theory and signal processing for communications.

Ron Dabora (M'07–SM'14) received the B.Sc. and M.Sc. degrees from Tel-Aviv University in 1994 and 2000, respectively, and the Ph.D. degree from Cornell University in 2007, all in electrical engineering. From 1994 to 2000, he was with the Ministry of Defense of Israel. From 2000 to 2003, he was with the Algorithms Group, Millimetrix Broadband Networks, Israel, and from 2007 to 2009, he was a post-doctoral researcher with the Department of Electrical Engineering, Stanford University. Since 2009, he has been an assistant professor with the Department of Electrical and Computer Engineering, Ben-Gurion University, Israel. His research interests include network information theory, wireless communications, and power line communications. He served as a TPC Member in a number of international conferences, including WCNC, PIMRC, and ICC. From 2012 to 2014, he served as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He currently serves as a Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS.