Phase Noise in MIMO Systems: Bayesian Cramér–Rao Bounds and Soft-Input Estimation

Ali A. Nasir, *Student Member, IEEE*, Hani Mehrpouyan, *Member, IEEE*, Robert Schober, *Fellow, IEEE*, and Yingbo Hua, *Fellow, IEEE*

Abstract—This paper addresses the problem of estimating time varying phase noise caused by imperfect oscillators in *multiple-input multiple-output* (MIMO) systems. The estimation problem is parameterized in detail and based on an equivalent signal model its dimensionality is reduced to minimize the overhead associated with phase noise estimation. New exact and closed-form expressions for the *Bayesian Cramér–Rao lower bounds* (BCRLBs) and soft-input *maximum a posteriori* (MAP) estimators for *online*, i.e., filtering, and *offline*, i.e., smoothing, estimation of phase noise over the length of a frame are derived. Simulations demonstrate that the proposed MAP estimators' *mean-square error* (MSE) performances are very close to the derived BCRLBs at moderate-to-high signal-to-noise ratios. To reduce the overhead and complexity associated with tracking the phase noise processes over the length of a frame, a novel soft-input *extended Kalman filter* (EKF) and *extended Kalman smoother* (EKS) that use soft statistics of the transmitted symbols given the current observations are proposed. Numerical results indicate that by employing the proposed phase tracking approach, the bit-error rate performance of a MIMO system affected by phase noise can be significantly improved. In addition, simulation results indicate that the proposed phase noise estimation scheme allows for application of higher order modulations and larger numbers of antennas in MIMO systems that employ imperfect oscillators.

Index Terms—Bayesian Cramér Rao lower bound (BCRLB), extended Kalman smoother (EKS), maximum-a-posteriori (MAP), multiple-input multiple-output (MIMO), soft-decision extended Kalman filter (EKF), Wiener phase noise.

I. INTRODUCTION

A. Motivation and Literature Survey

MULTIPLE-INPUT multiple-output (MIMO) technology can be utilized to enhance the throughput and reliability of wireless communication links by introducing multiplexing and diversity gains to the overall system [1], [2]. As a result, MIMO systems are an effective means to meet the stringent requirements on today's wireless communication systems that demand higher spectral efficiencies and throughputs [3]. On the other hand, phase noise severely deteriorates the performance of MIMO systems [4]–[12].

Oscillators used in bandpass communication systems suffer from phase instabilities, which are referred to as oscillator phase noise [4]–[10]. Phase noise is a time varying process that changes from symbol to symbol [13]–[15]. Moreover, the deteriorating effect of phase noise may be more severe in MIMO systems employing higher order modulations, given that in MIMO systems, independent oscillators may be used at each transmit and receive antenna resulting in multiple phase noise processes that need to be jointly estimated at the MIMO receiver [8], [10]–[12], [15]. The use of independent oscillators at each transmit and each receive antenna is well-motivated in applications where antennas need to be placed far apart from one another, e.g., in the case of line-of-sight (LoS) MIMO systems[1] [17]. As a result, even though Cramér-Rao lower bounds (CRLBs) and algorithms for estimation of phase noise in single-input single-output (SISO) systems have been extensively and thoroughly studied in [13], [14], [18]–[30], these results cannot be applied to the case of MIMO systems. Similarly, phase locked loops, that can be used in SISO systems for phase noise tracking, cannot be applied in LoS-MIMO systems where multiple phase noise parameters need to be tracked simultaneously at the receiver [10].

While in the case of single carrier systems, phase noise results in a rotation of the signal constellation, for multi carrier systems, e.g., orthogonal frequency division multiplexing (OFDM) systems, phase noise results in both a common phase noise error (CPE) and inter-carrier interference (ICI) [6], [31]–[39]. The CPE can be compensated by a phase rotation [6], [31], [33]. Moreover, extensive research has been carried out to mitigate the effect of ICI in OFDM systems [6], [32]. However, as outlined in [10], these algorithms are only applicable to SISO and MIMO systems where all transmit and receive antennas are equipped with a single oscillator. Thus, they cannot be applied for tracking multiple phase noise processes at the MIMO receiver. It is also important to note that compared to single carrier systems, phase noise is more detrimental to the performance of OFDM systems, especially as the constellation size and number of subcarriers increases [40]. Therefore, application of single carrier systems in very high speed communication links may be

[1]LoS MIMO has been demonstrated for microwave backhauling, e.g., by Ericsson AB [16].

advantageous. For example, in the case of high speed LoS microwave backhaul links that connect cellular base stations to the core network single carrier space division multiplexing (SDM) is used instead of OFDM [41], [42].

In [15], pilot-aided estimation of phase noise in single carrier MIMO systems is investigated, where a Wiener filter is applied to eliminate the ambiguity in the estimation problem and to obtain the phase noise values corresponding to each antenna. However, the scheme in [15] requires that during the transmission of a pilot symbol from one transmit antenna, the remaining antennas stay silent, i.e., the other antennas are off. As a result, this approach is bandwidth inefficient. In addition, [15] does not provide any means for finding lower bounds, e.g., a Bayesian Cramér-Rao lower bound (BCRLB), for phase noise estimation in MIMO systems. In [10], we address the problem of data-aided joint channel and phase noise estimation in MIMO systems. However, in [10], the BCRLB for the phase noise estimation problem is not derived and the proposed hard-decision feedback algorithm requires frequent transmission of pilot symbols.

B. Contributions

The results in this paper in combination with the joint channel and phase noise estimation algorithms in [10], provide a comprehensive approach for estimating channels and tracking phase noise parameters in MIMO systems. More specifically, by focusing on phase noise estimation, this paper seeks to improve phase noise tracking accuracy while reducing synchronization overhead. The contributions of this paper can be summarized as follows:

- Estimation of phase noise in MIMO systems is parameterized and an equivalent system model is derived. This signal model is applied to reduce the dimensionality and overhead associated with phase noise estimation.
- New exact expressions for data-aided (DA) BCRLBs, non-data-aided (NDA) BCRLBs, and soft-input maximum a posteriori (MAP) estimators for obtaining multiple phase noise parameters over the length of a frame in *offline*, i.e., smoothing, and *online*, i.e., filtering, modes are derived.[2] Simulation results indicate that the mean-square error (MSE) performances of the proposed MAP estimators are very close to the derived BCRLBs at moderate-to-high signal-to-noise ratios (SNRs).
- New soft-input extended Kalman filter (EKF) and extended Kalman smoother (EKS) algorithms, that exploit the statistics of the transmitted symbols given the current observations, are proposed to track the time varying phase noise over a frame. A complexity analysis is carried out to show that compared to the proposed MAP estimators and the algorithm in [10], [15], the proposed soft-input EKF-EKS can significantly reduce the computational complexity associated with tracking multiple phase noise processes in a MIMO system.
- It is demonstrated through extensive simulations that by employing the proposed soft-input EKF and EKS, the bit-error rate (BER) performance of a MIMO system in the

presence of phase noise can be significantly improved. These results further indicate that the proposed phase noise estimation scheme enables the application of higher order modulations and larger numbers of antennas in MIMO systems.

C. Organization

The remainder of the paper is organized as follows: In Section II, the phase noise model and MIMO framework used throughout the paper are outlined. Section III derives the offline and online BCRLBs for phase noise estimation, and Section IV presents the novel soft-input MAP and EKF-EKS based phase noise estimators. Section V provides numerical and simulation results that examine the performance of MIMO systems in the presence of estimated channel coefficients and phase noise. Finally, Section VI concludes the paper and summarizes its key findings.

Notations

Superscripts $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^T$ denote the conjugate, the conjugate transpose, and the transpose operators, respectively. Bold face small letters, e.g., $\mathbf{x}$, are used for vectors, bold face capital letters, e.g., $\mathbf{X}$, are used for matrices, $[\mathbf{X}]_{x,y}$ represents the entry in row x and column y of $\mathbf{X}$, and $[\mathbf{X}]_{x:\bar{x},y:\bar{y}}$ is used to denote a sub-matrix consisting of rows x to $\bar{x}$ and columns y to $\bar{y}$ of $\mathbf{X}$. $\mathbf{I}_{X \times X}$, $\mathbf{0}_{X \times X}$, and $\mathbf{1}_{X \times X}$ denote the $X \times X$ identity, all zero, and all one matrices, respectively. $|\cdot|$ is the absolute value operator, $|\mathbf{x}|$ denotes the element-wise absolute value of vector $\mathbf{x}$, $\det(\mathbf{X})$ represents the determinant of matrix $\mathbf{X}$, and $\mathrm{diag}(\mathbf{x})$ is used to denote a diagonal matrix, where the diagonal elements are given by vector $\mathbf{x}$. $\mathbf{X} \succeq \bar{\mathbf{X}}$ indicates that matrix $(\mathbf{X} - \bar{\mathbf{X}})$ is positive semi-definite. $\mathbb{E}[\cdot]$ denotes the expected value of the argument, and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary parts of a complex quantity, respectively. $\mathcal{N}\left(\mu, \sigma^2\right)$ and $\mathcal{CN}\left(\mu, \sigma^2\right)$ denote real and complex Gaussian distributions with mean μ and variance σ^2, respectively. $p(x)$ is used to denote the probability distribution of x. Finally, $\Delta_{\mathbf{x}}^{\mathbf{x}} f(\cdot) \triangleq \frac{\partial}{\partial \mathbf{x}} \left[\frac{\partial}{\partial \mathbf{x}} f(\cdot)\right]^T$ is a Hessian operator that calculates the second order partial derivative of function f with respect to vector $\mathbf{x}$.

II. SYSTEM MODEL

In this paper, a point-to-point MIMO system with N_t transmit and N_r receive antennas is considered (see Fig. 1). Each transmit and receive antenna is assumed to be equipped with an independent oscillator. Frame based transmission is assumed where the transmitter successively transmits frames of K data symbols. Throughout this paper, indices $m = 1, \ldots, N_t$, $n = 1, \ldots, N_r$, and $k = 1, \ldots, K$ are used to denote transmit antennas, receive antennas, and the time index, respectively. In Fig. 1, $h_{n,m}$ is the channel coefficient from the mth transmit to the nth receive antenna, which is assumed to be *known* and constant over the length of a frame, i.e., quasi-static fading channels are considered. The assumption of known channel gains is justified since the topic of joint channel and phase noise estimation using a *known training sequence* that is transmitted prior to data transmission is addressed in [10]. Therefore, in this paper, the main focus is on tracking the time varying
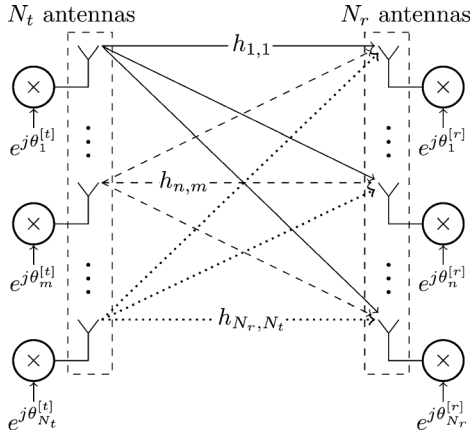
---

[2]The proposed soft-input estimators do not require the transmission of pilot symbols.

Fig. 1. System model for a point-to-point MIMO system.

phase noise using the transmitted data symbols[3]. Note that the assumption of quasi-static fading is in line with previous phase noise estimation algorithms in SISO and MIMO systems in [13]–[15], [18], [43]–[45]. Moreover, this assumption is reasonable in many practical scenarios, e.g., in LoS MIMO systems applied in microwave backhaul and satellite communication links [42], where the channel gains vary much more slowly than the phase noise process [39].

The vector of discrete-time baseband received signals at the N_r antennas of the MIMO receiver at time k, $\mathbf{y}(k) \triangleq [y_1(k), \ldots, y_{N_r}(k)]^T$, is given by [10]

$$\mathbf{y}(k) = \boldsymbol{\Theta}^{[r]}(k)\mathbf{H}\boldsymbol{\Theta}^{[t]}(k)\mathbf{s}(k) + \mathbf{w}(k), \quad k = 1, \ldots, K, \quad (1)$$

where

- $\mathbf{s}(k) \triangleq [s_1(k), \ldots, s_{N_t}(k)]^T$ with $s_m(k)$ denoting the M-ary modulated data symbol transmitted from the mth transmit antenna at time k and belonging to an alphabet $\mathcal{A}$ of size M,
- $\boldsymbol{\Theta}^{[r]}(k) \triangleq \mathrm{diag}\left(e^{j\theta_1^{[r]}(k)}, \ldots, e^{j\theta_{N_r}^{[r]}(k)}\right)$ is an $N_r \times N_r$ diagonal matrix, $\boldsymbol{\Theta}^{[t]}(k) \triangleq \mathrm{diag}\left(e^{j\theta_1^{[t]}(k)}, \ldots, e^{j\theta_{N_t}^{[t]}(k)}\right)$ is an $N_t \times N_t$ diagonal matrix, $\theta_n^{[r]}(k)$ and $\theta_m^{[t]}(k)$ correspond to the kth sample of the phase noise process at the nth receive and the mth transmit antennas, respectively,
- $\mathbf{H}$ is the *known* $N_r \times N_t$ channel matrix whose nth row and mth column entry, $h_{n,m}$, is assumed to be constant over the length of a frame but to change independently from frame to frame according to a complex Gaussian distribution, i.e., $h_{n,m} \sim \mathcal{CN}\left(\mu_{h_{n,m}}, \sigma_{h_{n,m}}^2\right)$, and
- $\mathbf{w}(k) \triangleq [w_1(k), \ldots, w_{N_r}(k)]^T$ with $w_n(k)$ denoting the zero-mean complex additive white Gaussian noise (AWGN) at the nth receive antenna, i.e., $w_n(k) \sim \mathcal{CN}\left(0, \sigma_w^2\right)$.

Throughout this paper the AWGN variance, σ_w^2, is assumed to be known since it can be estimated at the receiver [46]. Furthermore, transmitted symbols at different time instances are assumed to be statistically independent (uncoded system). By

---

[3]In Section V, the effect of *imperfectly* estimated channels on the performance of the proposed estimators and the overall MIMO system is thoroughly investigated.

arbitrarily selecting $\theta_{N_t}^{[t]}(k)$ as the reference phase value at time k, (1) can be written as

$$\mathbf{y}(k) = e^{j\theta_{N_t}^{[t]}(k)}\boldsymbol{\Theta}^{[r]}(k)\mathbf{H}\boldsymbol{\Theta}^{[t]}(k)\mathbf{s}(k)e^{-j\theta_{N_t}^{[t]}(k)} + \mathbf{w}(k)$$
$$= \boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{s}(k) + \mathbf{w}(k), \quad k = 1, \ldots, K, \quad (2)$$

where $\boldsymbol{\Phi}^{[r]}(k) \triangleq \mathrm{diag}\left(e^{j\phi_1^{[r]}(k)}, \ldots, e^{j\phi_{N_r}^{[r]}(k)}\right)$ and $\boldsymbol{\Phi}^{[t]}(k) \triangleq \mathrm{diag}\left(e^{j\phi_1^{[t]}(k)}, \ldots, e^{j\phi_{N_t-1}^{[t]}(k)}, 1\right)$ are $N_r \times N_r$ and $N_t \times N_t$ diagonal matrices, respectively, and $\phi_m^{[t]}(k) \triangleq \theta_m^{[t]}(k) - \theta_{N_t}^{[t]}(k)$ and $\phi_n^{[r]}(k) \triangleq \theta_n^{[r]}(k) + \theta_{N_t}^{[t]}(k)$ correspond to the modified phase noise processes at the kth sample at the mth transmit and nth receive antennas, respectively.

It can be observed that the system models in (1) and (2) are equivalent, where (1) describes the received signal at time k, $\mathbf{y}(k)$, in terms of $N_t + N_r$ independent phase noise values, $\boldsymbol{\Theta}(k) \triangleq \left[\theta_1^{[t]}(k), \ldots, \theta_{N_t}^{[t]}(k), \theta_1^{[r]}(k), \ldots, \theta_{N_r}^{[r]}(k)\right]^T$, and (2) in terms of $N \triangleq N_t + N_r - 1$ correlated phase noise values, $\boldsymbol{\phi}(k) \triangleq \left[\phi_1^{[t]}(k), \ldots, \phi_{N_t-1}^{[t]}(k), \phi_1^{[r]}(k), \ldots, \phi_{N_r}^{[r]}(k)\right]^T$, since $\phi_{N_t}^{[t]}(k) \triangleq \theta_{N_t}^{[t]}(k) - \theta_{N_t}^{[t]}(k) = 0$. The new parameterization in (2) eliminates the phase ambiguity, i.e., lack of phase reference, that is associated with the estimation of the $N_t + N_r$ independent phase noise parameters in (1) [15]. In fact, due to this phase ambiguity, it can be shown that the *Fisher's information matrix* for estimation of these $N_t + N_r$ independent phase noise parameters is singular. According to [47], a singular Fisher's information matrix indicates that there is no unbiased estimator for estimating these phase noise parameters with finite estimation error variance. However, the proposed equivalent signal model in (2) eliminates this phase ambiguity while reducing the number of parameters that need to be estimated. We note that the signal model in (2) assumes perfect timing and frame synchronization, which can be achieved by standard synchronization algorithms [13]–[15], [18], [43]–[45].

The discrete time phase noise model in (2) is motivated by the results in [6], [15], [18]. More importantly, for free-running oscillators, it is found that the phase noise process can be modeled as a Wiener process [4], [6], [7], [48]–[50]. Therefore, $\theta_m^{[t]}$ and $\theta_n^{[r]}$, for $m = 1, \ldots, N_t$ and $n = 1, \ldots, N_r$, are given by [4], [6], [7], [48]–[50]

$$\theta_m^{[t]}(k) = \theta_m^{[t]}(k-1) + \delta_m^{[t]}(k),$$
$$\theta_n^{[r]}(k) = \theta_n^{[r]}(k-1) + \delta_n^{[r]}(k), \quad (3)$$

where the phase innovations for the mth transmit and nth receive antennas, $\delta_m^{[t]}(k)$ and $\delta_n^{[r]}(k)$, respectively, are assumed to be real white Gaussian processes with $\delta_m^{[t]}(k) \sim \mathcal{N}\left(0, \sigma_{\delta_m^{[t]}}^2\right)$ and $\delta_n^{[r]}(k) \sim \mathcal{N}\left(0, \sigma_{\delta_n^{[r]}}^2\right)$. Since there is a direct relationship between the physical properties of the oscillators and the innovation variances at the mth transmit and nth receive antennas, $\sigma_{\delta_m^{[t]}}^2$ and $\sigma_{\delta_n^{[r]}}^2$, respectively [4], $\sigma_{\delta_m^{[t]}}^2$ and $\sigma_{\delta_n^{[r]}}^2$, $\forall\, m, n$, are assumed to be known at the receiver. Using (1) and (2), $\phi_m^{[t]}$ for $m = 1, \ldots, N_t - 1$ and $\phi_n^{[r]}$ for $n = 1, \ldots, N_r$ are given by

$$\phi_m^{[t]}(k) = \phi_m^{[t]}(k-1) + \eta_m^{[t]}(k),$$
$$\phi_n^{[r]}(k) = \phi_n^{[r]}(k-1) + \eta_n^{[r]}(k), \quad (4)$$

where $\eta_m^{[t]}(k) \triangleq \delta_m^{[t]}(k) - \delta_{N_t}^{[t]}(k)$ and $\eta_n^{[r]}(k) \triangleq \delta_n^{[r]}(k) + \delta_{N_t}^{[t]}(k)$ are real white Gaussian processes with $\eta_m^{[t]}(k) \sim \mathcal{N}\left(0, \sigma_{\delta_m^{[t]}}^2 + \sigma_{\delta_{N_t}^{[t]}}^2\right)$ and $\eta_n^{[r]}(k) \sim \mathcal{N}\left(0, \sigma_{\delta_n^{[r]}}^2 + \sigma_{\delta_{N_t}^{[t]}}^2\right)$.

III. BAYESIAN CRAMÉR-RAO LOWER BOUNDS

In order to find benchmarks for the performance of phase noise estimators over the length of a frame, in this section, we derive new exact and closed-form expressions for the DA-BCRLB for *offline* and *online* estimation of time varying phase noise in an $N_r \times N_t$ MIMO system. Since in most practical scenarios the transmitted *data symbols*, $\mathbf{s}(k)$, for $k = 1, \ldots, K$, are not known at the receiver, the derived data-aided bounds represent the highest possible estimation accuracy that can be achieved for phase noise estimation at a MIMO receiver based on the signal model in (2). Furthermore, an expression for the BCRLB for NDA estimation of phase noise is also obtained and numerically evaluated.

A. Data-Aided BCRLB for Offline Estimation

This subsection describes the detailed derivation of the DA-BCRLB for *offline* or smoothing estimation of KN phase noise processes, $\boldsymbol{\phi} \triangleq \left[\boldsymbol{\phi}^T(1), \ldots, \boldsymbol{\phi}^T(K)\right]^T$, in a frame of K received signals, $\mathbf{y} \triangleq \left[\mathbf{y}^T(1), \ldots, \mathbf{y}^T(K)\right]^T$, given $\mathbf{s} \triangleq \left[\mathbf{s}^T(1), \ldots, \mathbf{s}^T(K)\right]^T$. For the offline BCRLB, the whole frame of K received signals, $\mathbf{y}$, is assumed to be available at the receiver and used to estimate the corresponding KN phase noise values, where $N = N_t + N_r - 1$. The accuracy of the estimation of $\boldsymbol{\phi}$ is lower bounded by the offline DA-BCRLB as [51], [52]

$$\mathbb{E}_{\mathbf{y},\boldsymbol{\phi}|\mathbf{s}}\left[\left(\hat{\boldsymbol{\phi}}(\mathbf{y},\mathbf{s}) - \boldsymbol{\phi}\right)\left(\hat{\boldsymbol{\phi}}(\mathbf{y},\mathbf{s}) - \boldsymbol{\phi}\right)^T\right] \succeq \mathbf{B}^{-1}, \quad (5)$$

where $\hat{\boldsymbol{\phi}}(\mathbf{y},\mathbf{s})$ is the vector of estimated phase noise processes and $\mathbf{B}$ denotes the offline DA *Bayesian information matrix* (BIM) that is given by [52, p. 84],

$$\mathbf{B} = \mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})] + \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi})\right]. \quad (6)$$

In (6), $\mathbf{F}(\boldsymbol{\phi},\mathbf{s}) \triangleq \mathbb{E}_{\mathbf{y}|\boldsymbol{\phi},\mathbf{s}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\mathbf{y}|\boldsymbol{\phi},\mathbf{s})\right]$ is the DA-*Fisher's information matrix* (FIM) [46], $\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\mathbf{y} \mid \boldsymbol{\phi},\mathbf{s})$ is the Hessian of the log-likelihood function given $\boldsymbol{\phi}$ and $\mathbf{s}$, and

$\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi})\right]$ represents the information that is captured in the *probability density function* (PDF) of $\boldsymbol{\phi}$, i.e., $p(\boldsymbol{\phi})$.

Theorem: The BIM, $\mathbf{B}$ in (6), is a $KN \times KN$ symmetric block tridiagonal matrix that is given by

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\Xi}_1 & -\boldsymbol{\Sigma}^{-1} & 0 & \ldots & 0 \\ -\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Xi}_2 & -\boldsymbol{\Sigma}^{-1} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Xi}_{K-1} & -\boldsymbol{\Sigma}^{-1} \\ 0 & \ldots & 0 & -\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Xi}_K \end{bmatrix}, \quad (7)$$

where

$$\boldsymbol{\Xi}_k = \begin{cases} \mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})]_{k,k} + \boldsymbol{\Sigma}^{-1}, & k = 1, K \\ \mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})]_{k,k} + 2\boldsymbol{\Sigma}^{-1}, & k = 2, \ldots, K-1 \end{cases} \quad (8)$$

is an $N \times N$ matrix,
- $p(\mathbf{y}|\boldsymbol{\phi},\mathbf{s})$ is the log-likelihood function given $\mathbf{s}$,
- $\mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})]_{k,k}$, $k = 1, \ldots, K$, are the $N \times N$ diagonal blocks of the $KN \times KN$ block diagonal matrix $\mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})] = \mathbb{E}_{\boldsymbol{\phi}}\left[\mathbb{E}_{\mathbf{y}|\boldsymbol{\phi},\mathbf{s}}\left\{-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\mathbf{y}|\boldsymbol{\phi},\mathbf{s})\right\}\right] = \mathbb{E}_{\mathbf{y},\boldsymbol{\phi}|\mathbf{s}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\mathbf{y}|\boldsymbol{\phi},\mathbf{s})\right]$, and
- $[\mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})]_{k,k}]_{\ell,\bar{\ell}} = \left[\mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\phi}|\mathbf{s}}\left[-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)} \log p(\mathbf{y}(k)|\boldsymbol{\phi}(k), \mathbf{s}(k))\right]\right]_{\ell,\bar{\ell}}$, for $\bar{\mathbf{y}} \triangleq \left[\mathbf{y}^T(1), \ldots, \mathbf{y}^T(k)\right]^T$, can be determined as shown in (9) at the bottom of this page.

The $N \times N$ covariance matrix of $\boldsymbol{\phi}(k)$ given $\boldsymbol{\phi}(k-1)$, $\forall k$, $\boldsymbol{\Sigma}_{\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)}$, and its inverse can also be determined as (for notational simplicity, $\boldsymbol{\Sigma}_{\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)}$ is denoted by $\boldsymbol{\Sigma}$ throughout this paper)

$$\boldsymbol{\Sigma} = \mathbb{E}\left[[\boldsymbol{\phi}(k) - \boldsymbol{\phi}(k-1)][\boldsymbol{\phi}(k) - \boldsymbol{\phi}(k-1)]^H\right]$$
$$= \text{diag}(\sigma_{\delta_1^{[t]}}^2, \ldots, \sigma_{\delta_{N_t-1}^{[t]}}^2, \sigma_{\delta_1^{[r]}}^2, \ldots, \sigma_{\delta_{N_r}^{[r]}}^2) + \sigma_{\delta_{N_t}^{[t]}}^2 \mathbf{1}_{N \times N}, \quad (10)$$

$$\boldsymbol{\Sigma}^{-1} = \left(\mathbb{E}\left[[\boldsymbol{\phi}(k) - \boldsymbol{\phi}(k-1)][\boldsymbol{\phi}(k) - \boldsymbol{\phi}(k-1)]^H\right]\right)^{-1}$$
$$= \mathbf{U} - \frac{\sigma_{\delta_{N_t}^{[t]}}^2}{1+g} \mathbf{U}\mathbf{1}_{N \times N}\mathbf{U}. \quad (11)$$

In (11),

$$\mathbf{U} = \text{diag}\left(\frac{1}{\sigma_{\delta_1^{[t]}}^2}, \ldots, \frac{1}{\sigma_{\delta_{N_t-1}^{[t]}}^2}, \frac{1}{\sigma_{\delta_1^{[r]}}^2}, \ldots, \frac{1}{\sigma_{\delta_{N_r}^{[r]}}^2}\right)$$

$$[\mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})]_{k,k}]_{\ell,\bar{\ell}} = \begin{cases} \frac{2}{\sigma_w^2}|s_l(k)|^2 \sum_{n=1}^{N_r} |h_{n,\ell}|^2, & \ell = \bar{\ell} = 1, \ldots, N_t - 1 \\ \frac{2}{\sigma_w^2} \sum_{m=1}^{N_t} |h_{b,m}|^2 |s_m(k)|^2, & \ell = \bar{\ell} = N_t, \ldots, N, b = \ell - (N_t - 1) \\ \frac{2}{\sigma_w^2}|h_{\bar{\ell}-(N_t-1),l}|^2 |s_l(k)|^2, & \ell = 1, \ldots, N_t - 1, \bar{\ell} = N_t, \ldots, N, \\ \frac{2}{\sigma_w^2}|h_{\ell-(N_t-1),\bar{l}}|^2 |s_{\bar{l}}(k)|^2, & \ell = N_t, \ldots, N, \bar{\ell} = 1, \ldots, N_t - 1, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

and

$$g = \sigma^2_{\delta^{[t]}_{N_t}} \left(\sum_{m=1}^{N_t-1} \frac{1}{\sigma^2_{\delta^{[t]}_m}} + \sum_{n=1}^{N_r} \frac{1}{\sigma^2_{\delta^{[r]}_n}} \right).$$

Proof: See Appendix A.

The diagonal elements of $\mathbf{B}^{-1}$ provide the BCRLB for offline DA estimation of $\boldsymbol{\phi}$. Let $\mathbf{B}^{-1}_i$ denote the ith $N \times N$ diagonal block matrix of $\mathbf{B}^{-1}$. Using the matrix inversion proposition for symmetric block tridiagonal matrices in [53, Proposition 3.2], the block diagonal elements of, $\mathbf{B}^{-1}$, $\mathbf{B}^{-1}_i$, $i = 1, \ldots, K$, can be calculated as

$$\mathbf{B}^{-1}_i = \mathbf{P}_i \mathbf{Q}_i, \tag{12}$$

where $\mathbf{P}_i$ and $\mathbf{Q}_i$ are $N \times N$ matrices, given by

$$\mathbf{P}_i = \begin{cases} \mathbf{I}_{N \times N}, & i = 1 \\ \boldsymbol{\Sigma}\boldsymbol{\Xi}_1, & i = 2 \\ \boldsymbol{\Sigma}\boldsymbol{\Xi}_{i-1}\mathbf{P}_{i-1} - \mathbf{P}_{i-2}, & i = 3, \ldots, K \end{cases}$$

$$\mathbf{Q}_i = \begin{cases} \left(\boldsymbol{\Xi}_K \mathbf{P}_K - \boldsymbol{\Sigma}^{-1}\mathbf{P}_{K-1}\right)^{-1}, & i = K \\ \mathbf{Q}_K \boldsymbol{\Xi}_K \boldsymbol{\Sigma}, & i = K-1 \\ \mathbf{Q}_{i+1}\boldsymbol{\Xi}_{i+1}\boldsymbol{\Sigma} - \mathbf{Q}_{i+2}, & i = K-2, \ldots, 1. \end{cases}$$

By applying (12), the DA-BCRLB can be evaluated by finding the inverse of only an $N \times N$ matrix, $\left(\boldsymbol{\Xi}_K \mathbf{P}_K - \boldsymbol{\Sigma}^{-1}\mathbf{P}_{K-1}\right)$, instead of a $KN \times KN$ matrix, $\mathbf{B}$.

B. Data-Aided BCRLB for Online Estimation

This section evaluates the DA-BCRLB for online estimation of $\boldsymbol{\phi}(k)$ using only the past and current observations, $\bar{\mathbf{y}} \triangleq \left[\mathbf{y}^T(1), \ldots, \mathbf{y}^T(k)\right]^T$. It has been shown in [54] that the $N \times N$ online BIM at time index k, $\bar{\mathbf{B}}(k)$, can be obtained from the online BIM at time index $k-1$, $\bar{\mathbf{B}}(k-1)$, via the following recursion

$$\bar{\mathbf{B}}(k) = \mathbf{D}^{22}(k-1) - \mathbf{D}^{21}(k-1)$$
$$\times \left[\bar{\mathbf{B}}(k-1) + \mathbf{D}^{11}(k-1)\right]^{-1}\mathbf{D}^{12}(k-1), \tag{13}$$

where
- $\mathbf{D}^{11}(k-1) \triangleq \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta^{\boldsymbol{\phi}(k-1)}_{\boldsymbol{\phi}(k-1)} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big)\right] = \boldsymbol{\Sigma}^{-1}$,
- $\mathbf{D}^{12}(k-1) \triangleq \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta^{\boldsymbol{\phi}(k)}_{\boldsymbol{\phi}(k-1)} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big)\right] = -\boldsymbol{\Sigma}^{-1}$,
- $\mathbf{D}^{21}(k-1) \triangleq \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta^{\boldsymbol{\phi}(k-1)}_{\boldsymbol{\phi}(k)} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big)\right] = -\boldsymbol{\Sigma}^{-1}$, and
- $\mathbf{D}^{22}(k-1) \triangleq \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta^{\boldsymbol{\phi}(k)}_{\boldsymbol{\phi}(k)} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big)\right]$
$+\mathbb{E}_{\bar{\mathbf{y}}, \boldsymbol{\phi}|\mathbf{s}}\left[-\Delta^{\boldsymbol{\phi}(k)}_{\boldsymbol{\phi}(k)} \log p\big(\mathbf{y}(k)|\boldsymbol{\phi}(k), \mathbf{s}(k)\big)\right] = \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Pi}(k)$

In addition,

$$\boldsymbol{\Pi}(k) \triangleq \mathbb{E}_{\boldsymbol{\phi}}\left[\bar{\mathbf{F}}\big(\boldsymbol{\phi}(k), \mathbf{s}(k)\big)\right]$$
$$= \mathbb{E}_{\bar{\mathbf{y}}, \boldsymbol{\phi}|\mathbf{s}}\left[-\Delta^{\boldsymbol{\phi}(k)}_{\boldsymbol{\phi}(k)} \log p\big(\mathbf{y}(k)|\boldsymbol{\phi}(k), \mathbf{s}(k)\big)\right]$$

can be evaluated via (9), where $\bar{\mathbf{F}}\big(\boldsymbol{\phi}(k), \mathbf{s}(k)\big)$ is the online DA-FIM at time index k for estimating $\boldsymbol{\phi}(k)$. Assuming no prior information about $\boldsymbol{\phi}(0)$, the initial online BIM, $\bar{\mathbf{B}}(0)$, is determined as [55]

$$\bar{\mathbf{B}}(0) = \mathbb{E}_{\boldsymbol{\phi}(0)}\left[-\Delta^{\boldsymbol{\phi}(0)}_{\boldsymbol{\phi}(0)} \log p\big(\boldsymbol{\phi}(0)\big)\right] = \mathbf{0}_{N \times N}.$$

Subsequently, the DA-BCRLB for estimating $\boldsymbol{\phi}(k)$ is given by

$$\bar{\mathbf{B}}^{-1}(k) = \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Pi}(k)\right.$$
$$\left. -\boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{B}}(k-1) + \boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1}\right)^{-1}. \tag{14}$$

C. Non-Data-Aided BCRLB

The NDA-BCRLB assumes no knowledge of the transmitted symbols, $\mathbf{s}(k)$, $\forall\, k$, and can be derived by taking into account the a priori distribution of the transmitted symbols. Thus, to evaluate the NDA-BCRLB, the likelihood function $p(\mathbf{y}|\boldsymbol{\phi})$ is obtained by averaging out the $p(\mathbf{y}|\boldsymbol{\phi}, \mathbf{s})$, in (A.1), over the a priori distribution of $\mathbf{s}$. Accordingly, assuming all elements of $\mathcal{A}$ are equiprobable, the log likelihood function, $\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k))$, is given by

$$\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k)) = \log \sum_{i=0}^{S-1} \frac{1}{S(\pi\sigma^2_w)^{N_r}}$$
$$\times \exp\left\{-\frac{\left\|\mathbf{y}(k) - \boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{c}_i\right\|^2}{\sigma^2_w}\right\}, \tag{15}$$

where $\mathbf{c}_i$, for, $i = 0, \ldots, S-1$, denotes the $N_t \times 1$ vector of the ith permutation of all possible symbols belonging to alphabet $\mathcal{A}$ and $S = M^{N_t}$. It is clear from (15) that the summation inside the $\log$ function cannot be further simplified analytically. However, following the procedure given in (A.3)–(A.9) for the DA-BCLRB, we can take the second derivative of $-\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k))$, in (15), with respect to the phase noise parameters. Consequently, the expected value of the NDA-FIM, $\mathcal{F}(\boldsymbol{\phi}) \triangleq \mathbb{E}_{\mathbf{y}|\boldsymbol{\phi}}\left[-\Delta^{\boldsymbol{\phi}}_{\boldsymbol{\phi}} \log p(\mathbf{y}|\boldsymbol{\phi})\right]$, with respect to $\boldsymbol{\phi}$ is given by (16) at the bottom of the next page. In (16)

- $\zeta_m \triangleq \Re\left\{\mathbf{y}^H(k)\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\ddot{\boldsymbol{\Phi}}^{[t]}_m(k)\mathbf{c}_i - \mathbf{c}^H_i\big(\boldsymbol{\Phi}^{[t]}(k)\big)^H \mathbf{H}^H\mathbf{H}\dot{\boldsymbol{\Phi}}^{[t]}_m(k)\mathbf{c}_i\right\}$,
- $\zeta_n \triangleq \Re\left\{\mathbf{y}^H(k)\dot{\boldsymbol{\Phi}}^{[r]}_n(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{c}_i\right\}$
- $\zeta_{m,m} \triangleq \Re\left\{\mathbf{y}^H(k)\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\ddot{\boldsymbol{\Phi}}^{[t]}_m(k)\mathbf{c}_i - \mathbf{c}^H_i\big(\boldsymbol{\Phi}^{[t]}(k)\big)^H \mathbf{H}^H\mathbf{H}\ddot{\boldsymbol{\Phi}}^{[t]}_m(k)\mathbf{c}_i - \mathbf{c}^H_i\big(\dot{\boldsymbol{\Phi}}^{[t]}_m(k)\big)^H\mathbf{H}^H\mathbf{H}\dot{\boldsymbol{\Phi}}^{[t]}_m(k)\mathbf{c}_i\right\}$,
- $\zeta_{n,n} \triangleq \Re\left\{\mathbf{y}^H(k)\ddot{\boldsymbol{\Phi}}^{[r]}_n(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{c}_i\right\}$,
- $\zeta_{m,n} \triangleq \Re\left\{\mathbf{y}^H(k)\dot{\boldsymbol{\Phi}}^{[r]}_n(k)\mathbf{H}\dot{\boldsymbol{\Phi}}^{[t]}(k)\mathbf{c}_i\right\}$,
- $\zeta_{m,\bar{m}} \triangleq \Re\left\{-\mathbf{c}^H_i\big(\dot{\boldsymbol{\Phi}}^{[t]}_{\bar{m}}(k)\big)^H\mathbf{H}^H\mathbf{H}\dot{\boldsymbol{\Phi}}^{[t]}_m(k)\mathbf{c}_i\right\}$ for $m, \bar{m} = 1, \ldots, N_t - 1$ and $n = 1, \ldots, N_r$ depend on $\mathbf{c}_i$, for, $i = 0, \ldots, S-1$,
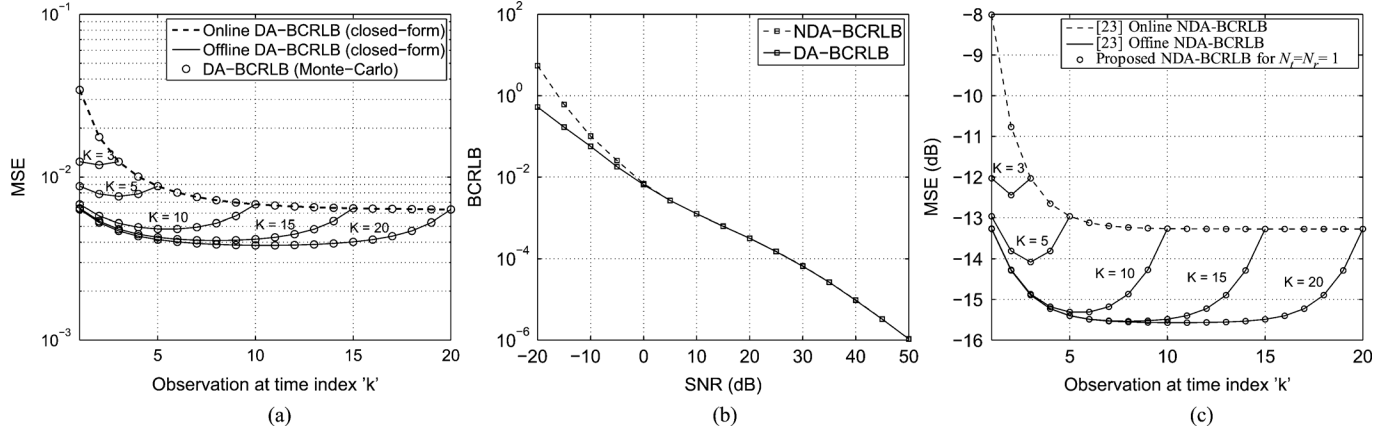- $\alpha \triangleq \frac{2}{S\sigma^2_w(\pi\sigma^2_w)^{N_r}}$,

Fig. 2. (a) Online and offline DA-BCRLB versus the number of observations for 2×2 MIMO system with $\mathrm{SNR} = 5$ dB and $\sigma_\delta^2 = 10^{-3}\ \mathrm{rad}^2$. (b) NDA-and DA-BCRLB for 2×2 MIMO system versus SNR evaluated at $k = 20$ in a frame of $K = 20$ symbols with $\sigma_\delta^2 = 10^{-4}\ \mathrm{rad}^2$. (c) Comparison of derived online and offline NDA-BCRLB results with [23] for $\mathrm{SNR} = 5$ dB, $N_t = N_r = 1$, and $\sigma_\delta^2 = 10^{-2}\ \mathrm{rad}^2$.

- $\beta \triangleq \exp\left\{ -\frac{\left\| \mathbf{y}(k) - \boldsymbol{\Phi}^{[r]}(k) \mathbf{H} \boldsymbol{\Phi}^{[t]}(k) \mathbf{c}_i \right\|^2}{\sigma_w^2} \right\}$,

- $\nu \triangleq \sum_{i=0}^{S-1} \frac{1}{S(\pi\sigma_w^2)^{N_r}} \exp\left\{ -\frac{\left\| \mathbf{y}(k) - \boldsymbol{\Phi}^{[r]}(k) \mathbf{H} \boldsymbol{\Phi}^{[t]}(k) \mathbf{c}_i \right\|^2}{\sigma_w^2} \right\}$,
and
- $\dot{\boldsymbol{\Phi}}_n^{[r]}(k)$, $\ddot{\boldsymbol{\Phi}}_n^{[r]}(k)$, $\dot{\boldsymbol{\Phi}}_m^{[t]}(k)$, and $\ddot{\boldsymbol{\Phi}}_m^{[t]}(k)$ are defined in Appendix A.

Unlike the DA-BCRLB, the expectation in (16) over $\mathbf{y}(k)$ and $\boldsymbol{\phi}(k)$ cannot be analytically evaluated and does not result in a closed-form expression. Thus, to evaluate the offline and online NDA-BCRLBs, the NDA-FIM in (16) needs to be applied instead of $\mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi}, \mathbf{s})]$ in (12) and (14), respectively, and the resulting expectation is numerically evaluated.

D. Numerical Evaluation of the BCRLB

Fig. 2 plots the derived DA-and NDA-BCRLBs. It is assumed that the phase noise innovation variances are $\sigma_\delta^2 = \sigma_{\delta_m^{[t]}}^2 = \sigma_{\delta_n^{[r]}}^2, \forall\, m, n$. Binary phase-shift keying (BPSK) modulation is used for the data symbols, $\mathbf{s}(k)$. A specific channel, $\mathbf{H} = [0.9928 + 0.2920i, -0.6541 - 1.2625i; 1.2740 - 0.2759i, 0.3207 - 2.0030i]$, is considered for the ease of comparison [56]. Finally, Fig. 2 only displays the BCRLB for the estimation of $\phi_{N_r}^{[r]}(k), \forall\, k$. Similar results are obtained for the remaining phase noise processes.

Fig. 2(a) plots the offline and online DA-BCRLBs for frame lengths of $K = [3, 5, 10, 15, 20]$ and a 2×2 MIMO system with an SNR of $1/\sigma_w^2 = 5$ dB. In order to verify the analytical results, the offline and online DA-BCRLBs are also evaluated using Monte-Carlo simulations and the results are plotted in Fig. 2(a), where the BIMs for both cases were obtained by evaluating the expectation of the Hessian matrix, $-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)} \log p\big(\mathbf{y}(k)|\boldsymbol{\phi}(k)\big)$, over $\mathbf{y}$ and $\boldsymbol{\phi}$, for 1000 Monte-Carlo simulations and $\boldsymbol{\phi}$ was generated according to the Wiener process in (3) in each simulation. The results in Fig. 2(a) validate the analysis in this section since the offline and online DA-BCRLBs evaluated using the closed-form expressions in (12) and (14) coincide with the Monte-Carlo simulation results. Fig. 2(b) plots the NDA-and DA-BCRLB for a 2×2 MIMO system versus SNR ($k = 20$ in a frame of $K = 20$ symbols and $\sigma_\delta^2 = 10^{-4}\ \mathrm{rad}^2$).[4] Fig. 2(b) shows that the NDA-BCRLB approaches the DA-BCRLB at medium-to-high SNRs ($\mathrm{SNR} \geq 0$ dB). Finally, in Fig. 2(c), the derived online and offline NDA-BCRLBs are compared against the NDA bounds in [23] for SISO systems ($\mathrm{SNR} = 5$ dB, $N_t = N_r = 1$, and $\sigma_\delta^2 = 10^{-2}\ \mathrm{rad}^2$). Fig. 2(c) shows that for $N_t = N_r = 1$, the derived NDA-BCRLBs match with the results in [23].

The results in Fig. 2 also reveal that the minimum and maximum values of the offline BCRLB are achieved at the middle

[4]Since BCRLBs are plotted at $k = K = 20$, the offline and online results are identical as shown in Fig. 2(a).

$$
\left[\mathbb{E}_{\boldsymbol{\phi}}[\mathcal{F}(\boldsymbol{\phi})]_{k,k}\right]_{\ell,\bar{\ell}} = \mathbb{E}_{\mathbf{y}(k),\boldsymbol{\phi}(k)}
\begin{cases}
\frac{\alpha}{\nu}\sum_{i=0}^{S-1}(\beta\zeta_{\ell,\ell} + \frac{2}{\sigma_w^2}\zeta_\ell^2\beta) + \frac{\alpha^2}{\nu^2}\left(\sum_{i=0}^{S-1}\beta\zeta_\ell\right)^2, & \ell = \bar{\ell} = 1,\ldots,N_t-1 \\[6pt]
\frac{\alpha}{\nu}\sum_{i=0}^{S-1}(\beta\zeta_{b,b} + \frac{2}{\sigma_w^2}\zeta_b^2\beta) + \frac{\alpha^2}{\nu^2}\left(\sum_{i=0}^{S-1}\beta\zeta_b\right)^2, & \ell = \bar{\ell} = N_t,\ldots,N, b = \ell-(N_t-1) \\[6pt]
\frac{\alpha}{\nu}\sum_{i=0}^{S-1}(\beta\zeta_{\ell,b} + \frac{2}{\sigma_w^2}\zeta_\ell\zeta_b\beta) + \frac{\alpha^2}{\nu^2}\sum_{i=0}^{S-1}\beta\zeta_b\sum_{i=0}^{S-1}\beta\zeta_\ell, & \ell = 1,\ldots,N_t-1, \bar{\ell} = N_t,\ldots,N, b = \bar{\ell}-(N_t-1) \\[6pt]
\frac{\alpha}{\nu}\sum_{i=0}^{S-1}(\beta\zeta_{\bar{\ell},b} + \frac{2}{\sigma_w^2}\zeta_{\bar{\ell}}\zeta_b\beta) + \frac{\alpha^2}{\nu^2}\sum_{i=0}^{S-1}\beta\zeta_b\sum_{i=0}^{S-1}\beta\zeta_{\bar{\ell}}, & \ell = N_t,\ldots,N, \bar{\ell} = 1,\ldots,N_t-1, b = \ell-(N_t-1) \\[6pt]
\frac{\alpha}{\nu}\sum_{i=0}^{S-1}(\beta\zeta_{\ell,\bar{\ell}} + \frac{2}{\sigma_w^2}\zeta_\ell\zeta_{\bar{\ell}}\beta) + \frac{\alpha^2}{\nu^2}\sum_{i=0}^{S-1}\beta\zeta_{\bar{\ell}}\sum_{i=0}^{S-1}\beta\zeta_\ell, & \ell = 1,\ldots,N_t-1, \bar{\ell} = 1,\ldots,N_t-1, \ell \neq \bar{\ell} \\[6pt]
\frac{\alpha}{\nu}\sum_{i=0}^{S-1}(\frac{2}{\sigma_w^2}\zeta_{\bar{\ell}}\zeta_\ell\beta) + \frac{\alpha^2}{\nu^2}\sum_{i=0}^{S-1}\beta\zeta_\ell\sum_{i=0}^{S-1}\beta\zeta_{\bar{\ell}}, & \ell = N_t,\ldots,N, \bar{\ell} = N_t,\ldots,N, \ell \neq \bar{\ell}
\end{cases}
\tag{16}
$$

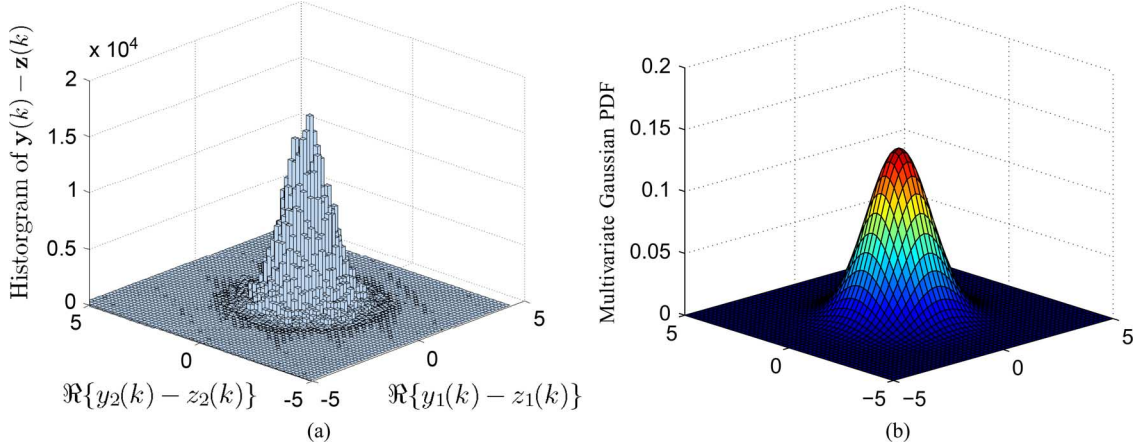Fig. 3. At $\mathrm{SNR} = 0$ dB and $N_t = N_r = 2$. (a) Histogram of $\mathbf{y}(k) - \mathbf{z}(k)$. (b) Multivariate Gaussian PDF with mean zero and covariance $\mathbf{\Sigma}_{\tilde{\mathbf{w}}(k)}\left(\boldsymbol{\phi}(k)\right)$.

and end-points of a frame, respectively, for any frame length, K. This implies that the best phase noise estimate can be achieved for the middle symbol within the frame, whereas the estimates get poorer as one moves to the boundary points. This behavior is expected since the phase noise for the symbol in the middle of the frame is followed by the largest number of past and future symbols with highly correlated phase noise values. Thus, by exploiting the observed symbols and correlations, the phase noise values corresponding to the middle symbol can be estimated with the highest accuracy. This can be observed from the time dependency of the BCRLB in (6), which is due to the prior information on the PDF of the phase noise processes. Moreover, Fig. 2 shows that the online BCRLB decreases with increasing observation length, K, since the longer the length of the observation sequence the more information is available for estimation of the kth symbol's phase noise. However, this gain in estimation accuracy with increasing K reaches an error floor since the overall estimation performance is ultimately limited by the AWGN regardless of the number of observation symbols in a frame.

IV. PHASE NOISE ESTIMATION

In this section, the soft-input MAP and soft-input EKF and EKS for tracking the phase noise processes are derived. Let us first establish the soft statistics of the transmitted symbols. Let $\mathbf{c}_i$, for, $i = 0, \ldots, S - 1$, with $S = M^{N_t}$, denote the $N_t \times 1$ vector of the ith permutation of all the possible symbols belonging to alphabet $\mathcal{A}$. Using Bayes' rule, the posterior probability $p\big(\mathbf{s}(k) = \mathbf{c}_i | \mathbf{y}(k), \boldsymbol{\phi}(k)\big)$ is given by

$$
\begin{aligned}
\rho_i(k) &\triangleq p\big(\mathbf{s}(k) = \mathbf{c}_i | \mathbf{y}(k), \boldsymbol{\phi}(k)\big) \\
&= \frac{p\big(\mathbf{y}(k) | \boldsymbol{\phi}(k), \mathbf{s}(k) = \mathbf{c}_i\big) p\big(\mathbf{s}(k) = \mathbf{c}_i\big)}{\sum_{i=0}^{S-1} p\big(\mathbf{y}(k) | \boldsymbol{\phi}(k), \mathbf{s}(k) = \mathbf{c}_i\big) p\big(\mathbf{s}(k) = \mathbf{c}_i\big)} \\
&= \frac{p\big(\mathbf{y}(k) | \boldsymbol{\phi}(k), \mathbf{s}(k) = \mathbf{c}_i\big)}{\sum_{i=0}^{S-1} p\big(\mathbf{y}(k) | \boldsymbol{\phi}(k), \mathbf{s}(k) = \mathbf{c}_i\big)}, \\
i &= 0, \ldots, S - 1,
\end{aligned}
\tag{17}
$$

where

$$
\begin{aligned}
&p\big(\mathbf{y}(k) | \boldsymbol{\phi}(k), \mathbf{s}(k) \\
&= \mathbf{c}_i\big) = \frac{1}{(\pi \sigma_w^2)^{N_r}} \exp\left\{ -\frac{\left\| \mathbf{y}(k) - \boldsymbol{\Phi}^{[r]}(k) \mathbf{H} \boldsymbol{\Phi}^{[t]}(k) \mathbf{c}_i \right\|^2}{\sigma_w^2} \right\}.
\end{aligned}
$$

It is assumed in (17) that $p(\boldsymbol{\phi}(k))$ and $p(\mathbf{s}(k) = \mathbf{c}_i)$ are independent random processes, i.e., $p\big(\boldsymbol{\phi}(k), \mathbf{s}(k) = \mathbf{c}_i\big) = p(\boldsymbol{\phi}(k)) p\big(\mathbf{s}(k) = \mathbf{c}_i\big)$ and all symbols have identical a priori probabilities, i.e., $p\big(\mathbf{s}(k) = \mathbf{c}_i\big) = 1/M^{N_t}, \forall i$. Let $\mathcal{P}(k) \triangleq [\rho_0(k), \ldots, \rho_{S-1}(k)]$ denote the *probability mass function* (PMF) of the kth transmitted symbol given the received signal. Given $\mathcal{P}(k)$, the vector of transmitted symbols, $\mathbf{s}(k)$, can be treated as a random signal and decomposed into two parts

$$
\mathbf{s}(k) = \bar{\mathbf{s}}(k) + \tilde{\mathbf{s}}(k),
\tag{18}
$$

where $\bar{\mathbf{s}}(k) \triangleq \mathbb{E}_{\mathcal{P}(k)}\big[\mathbf{s}(k)\big]$ is the mean of the data symbol at the kth sampling instant and $\tilde{\mathbf{s}}(k)$ denotes the random variation about $\bar{\mathbf{s}}(k)$ with the $N_t \times N_t$ covariance matrix, $\mathbf{\Sigma}_{\tilde{\mathbf{s}}(k)} = \mathbb{E}\big[\tilde{\mathbf{s}}(k)\tilde{\mathbf{s}}^H(k)\big] = \mathbb{E}_{\mathcal{P}(k)}\big[\mathbf{s}(k)\mathbf{s}^H(k)\big] - \bar{\mathbf{s}}(k)\bar{\mathbf{s}}^H(k)$. Since the distribution of $\tilde{\mathbf{s}}(k)$ is too complicated to be applied in the design of the soft-input estimators, here, $\tilde{\mathbf{s}}(k)$ is modelled as a zero-mean Gaussian random variable. The simulation results in Section V-A demonstrate that the performances of the soft-input MAP estimators derived based on this assumption are close to the BCRLB. We note that such an assumption/approximation has been widely applied in the literature, e.g., [57] and references therein. Moreover, using (18), the vector of received signals in (1) can be written as

$$
\mathbf{y}(k) = \underbrace{\boldsymbol{\Phi}^{[r]}(k) \mathbf{H} \boldsymbol{\Phi}^{[t]}(k) \bar{\mathbf{s}}(k)}_{\triangleq \mathbf{z}(k)} + \underbrace{\boldsymbol{\Phi}^{[r]}(k) \mathbf{H} \boldsymbol{\Phi}^{[t]}(k) \tilde{\mathbf{s}}(k) + \mathbf{w}(k)}_{\triangleq \tilde{\mathbf{w}}(k)},
\tag{19}
$$

where $\mathbf{z}(k)$ is the mean of the observation vector, $\mathbf{y}(k)$, and $\tilde{\mathbf{w}}(k)$ denotes the zero-mean random variation about $\mathbf{z}(k)$. Because of the Gaussian assumption, $\mathbf{y}(k)$ can be modeled as a

multivariate Gaussian random variable, with mean $\mathbf{z}(k)$ and $N_r \times N_r$ covariance matrix

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}\big(\boldsymbol{\phi}(k)\big) &= \mathbb{E}\left[\tilde{\mathbf{w}}(k)\tilde{\mathbf{w}}^H(k)\right] \\
&= \boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\boldsymbol{\Sigma}_{\tilde{\mathbf{s}}(k)}\big(\boldsymbol{\Phi}^{[t]}(k)\big)^H \\
&\quad \times \mathbf{H}^H\big(\boldsymbol{\Phi}^{[r]}(k)\big)^H + \sigma_w^2 \mathbf{I}_{N_r \times N_r}. \quad (20)
\end{aligned}
$$

Fig. 3(a) depicts the histogram of $\mathbf{y}(k)-\mathbf{z}(k)$ for 2×10^6 simulation trials, evaluated at SNR $=0$ dB and $N_t=N_r=2$.[5] Fig. 3(b) plots the bivariate Gaussian PDF evaluated with mean zero and covariance $\boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}\big(\boldsymbol{\phi}(k)\big)$, defined in (20), at SNR $=0$ dB. The results in Fig. 3(a) and (b) suggest that the PDF of $\mathbf{y}(k)-\mathbf{z}(k)$ can indeed be closely approximated by a Gaussian distribution.

Remark 1: By exploiting the soft statistics of the transmitted symbols as detailed above, the soft-input estimation and soft decoding are performed using the following steps:

Step 1: To perform *soft estimation* of phase noise processes at the kth instant, $\hat{\boldsymbol{\phi}}(k)$, the elements of the PMF in (17) are obtained based on the phase noise estimate at the $(k-1)$th time instant, i.e., $\boldsymbol{\phi}(k)=\hat{\boldsymbol{\phi}}(k-1)$.

Step 2: *Soft decoding* is carried out by evaluating $\rho_i(k)$ in (17), $\forall\,i$, using the phase noise estimate at the kth time instant, i.e., $\boldsymbol{\phi}(k)=\hat{\boldsymbol{\phi}}(k)$. Subsequently, the decoded symbol, $\hat{\mathbf{s}}(k)$, is given by the permutation, $\mathbf{c}_i$, that maximizes the posterior probability $\rho_i(k)$.

Step 3: *Steps 1* and *2* are repeated for all symbols in the frame.

A. Soft-Input MAP Phase Noise Estimation

In this subsection, offline and online soft-input MAP estimators for tracking phase noise over a frame are derived.

1) Offline MAP: The offline soft-input MAP estimator estimates KN phase noise values, $\boldsymbol{\phi}$, by employing all the K received data symbols in a frame, $\mathbf{y}$. Thus, the offline soft-input MAP estimate of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}^{[\mathrm{MAP}]}$, maximizes the log of the posterior PDF

$$
\begin{aligned}
\log p(\boldsymbol{\phi}|\mathbf{y}) &= \log\left\{\frac{p(\mathbf{y}|\boldsymbol{\phi})p(\boldsymbol{\phi})}{p(\mathbf{y})}\right\} \\
&= \sum_{k=1}^{K}\left\{\log\frac{p\big(\boldsymbol{\phi}(1)\big)}{(\pi\sigma_w^2)^{N_r}(2\pi)^{N/2}(\det\boldsymbol{\Sigma})^{1/2}}\right. \\
&\quad -\frac{[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]^H\boldsymbol{\Sigma}^{-1}[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]}{2} \\
&\quad -\left[\mathbf{y}(k)-\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\bar{\mathbf{s}}(k)\right]^H\boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}^{-1} \\
&\quad \left.\times\left[\mathbf{y}(k)-\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\bar{\mathbf{s}}(k)\right]\right\} \\
&\quad -\log p(\mathbf{y}). \quad (21)
\end{aligned}
$$

[5]The histogram of the real part of $\mathbf{y}(k)-\mathbf{z}(k)$ is plotted only because similar results are obtained for the imaginary part of $\mathbf{y}(k)-\mathbf{z}(k)$.

Since the first term inside the summation in (21) and $\log p(\mathbf{y})$ are independent of $\boldsymbol{\phi}$, the offline soft-input MAP estimate of $\boldsymbol{\phi}$ is given by

$$
\begin{aligned}
\hat{\boldsymbol{\phi}}^{[\mathrm{MAP}]} &= \arg\min_{\boldsymbol{\phi}}\sum_{k=1}^{K}\left\{\left[\mathbf{y}(k)-\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\bar{\mathbf{s}}(k)\right]^H\right. \\
&\quad \times \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}^{-1}\left[\mathbf{y}(k)-\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\bar{\mathbf{s}}(k)\right] \\
&\quad \left.+\frac{[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]^H\boldsymbol{\Sigma}^{-1}[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]}{2}\right\}.
\end{aligned}
$$
$$(22)$$

In order to reduce the computational complexity of the exhaustive search in (22), *alternating projection* (AP) can be used to reduce the KN-dimensional exhaustive search into a series of one-dimensional searches [58]. Note that AP is not guaranteed to converge to the true estimates and is dependent on the initialization. Thus, to achieve proper initialization, $\hat{\boldsymbol{\phi}}(k)$ is initialized by $\mathbf{0}_{N\times1}$, for $k=1,\ldots,K$, since the overall matrix of channel gains and phase noise processes is jointly estimated at the beginning of each frame [10].

2) Online MAP: The online soft-input MAP estimate of the phase noise vector at time k, $\hat{\boldsymbol{\phi}}^{[\mathrm{MAP}]}(k)$, maximizes the $\log$ of the posterior PDF at the kth symbol, i.e., $\log p\big(\mathbf{y}(k)|\boldsymbol{\phi}(k)\big)p\big(\boldsymbol{\phi}(k)|\hat{\boldsymbol{\phi}}(k-1)\big)$, and can be determined as

$$
\begin{aligned}
\hat{\boldsymbol{\phi}}^{[\mathrm{MAP}]}(k) &= \arg\min_{\boldsymbol{\phi}(k)}\left\{\left[\mathbf{y}(k)-\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\bar{\mathbf{s}}(k)\right]^H\right. \\
&\quad \times\boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}^{-1}\left[\mathbf{y}(k)-\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\bar{\mathbf{s}}(k)\right] \\
&\quad \left.+\frac{[\boldsymbol{\phi}(k)-\hat{\boldsymbol{\phi}}(k-1)]^H\boldsymbol{\Sigma}^{-1}[\boldsymbol{\phi}(k)-\hat{\boldsymbol{\phi}}(k-1)]}{2}\right\}.
\end{aligned}
$$
$$(23)$$

Similar to the offline MAP estimator, AP can be applied to reduce the N-dimensional exhaustive search in (23).

Remark 2: Due to their high computational complexities, the proposed soft-input MAP estimators may be difficult to implement in practice. Nevertheless, they are derived and applied in this paper to illustrate that there exist estimators with MSE performances that are very close to the derived BCRLBs. Thus, these tight bounds are validated and can also serve as bounds for practical estimators. Moreover, the derivation of a closed-form soft-input MAP estimator is topic of future research.

B. Soft-Input Extended Kalman Filter and Smoother

This subsection presents an EKF and EKS for estimating phase noise processes over the entire frame. Unlike the results in [10], where hard decisions of the previous symbols are used to obtain the current symbol's phase noise values, here, a soft-input EKF is derived to estimate current symbol's phase noise. Moreover, a new soft-input EKS that further refines the phase noise estimates over the frame is proposed.

The state and observation equation for the soft-input EKF are given by

$$
\boldsymbol{\phi}(k) = \boldsymbol{\phi}(k-1) + \boldsymbol{\eta}(k), \quad (24)
$$

and (19), respectively. In (24), $\boldsymbol{\eta}(k) \sim (\mathbf{0}_{N\times 1}, \boldsymbol{\Sigma})$. Since the observation in (19) is a non-linear function of the unknown state vector $\boldsymbol{\phi}(k)$, the EKF is used instead of a standard Kalman filter [46]. Thus, the $N_r \times N$ Jacobian matrix evaluated at $\boldsymbol{\phi}(k)$, $\mathbf{E}(\boldsymbol{\phi}(k))$, is determined by computing the first order partial derivative of $\mathbf{z}(k)$, defined in (19), with respect to the state vector $\boldsymbol{\phi}(k)$ as

$$\mathbf{E}(\boldsymbol{\phi}(k)) = \frac{\partial \mathbf{z}(k)}{\partial \boldsymbol{\phi}^T(k)} \tag{25}$$

$$= \left[\frac{\partial \mathbf{z}(k)}{\partial \phi_1^{[t]}(k)}, \dots, \frac{\partial \mathbf{z}(k)}{\partial \phi_{N_t-1}^{[t]}(k)}, \frac{\partial \mathbf{z}(k)}{\partial \phi_1^{[r]}(k)}, \dots, \frac{\partial \mathbf{z}(k)}{\partial \phi_{N_r}^{[r]}(k)} \right], \tag{26}$$

where $\frac{\partial \mathbf{z}(k)}{\partial \phi_m^{[t]}(k)} = \boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\bar{\mathbf{s}}(k)$, for $m = 1, \dots, N_t - 1$, $\frac{\partial \mathbf{z}(k)}{\partial \phi_n^{[r]}(k)} = \dot{\boldsymbol{\Phi}}_n^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\bar{\mathbf{s}}(k)$, for $n = 1, \dots, N_r$, and $\dot{\boldsymbol{\Phi}}_m^{[t]}(k)$ and $\dot{\boldsymbol{\Phi}}_n^{[r]}(k)$ are defined in Appendix A. Using (17)–(20) and (25), the remaining equations for the soft-input EKF are obtained as

$$\hat{\boldsymbol{\phi}}(k|k-1) = \hat{\boldsymbol{\phi}}(k-1|k-1), \tag{27}$$

$$\mathbf{M}(k|k-1) = \mathbf{M}(k-1|k-1) + \boldsymbol{\Sigma}, \tag{28}$$

$$\begin{aligned} \mathbf{G}(k) = \mathbf{M}(k|k-1)\mathbf{E}^H(\boldsymbol{\phi}(k|k-1)) \\ \times \Big(\mathbf{E}(\boldsymbol{\phi}(k|k-1))\mathbf{M}(k|k-1) \\ \times \mathbf{E}^H(\boldsymbol{\phi}(k|k-1)) \\ + \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}(\boldsymbol{\phi}(k|k-1)) \Big)^{-1}, \end{aligned} \tag{29}$$

$$\begin{aligned} \hat{\boldsymbol{\phi}}(k|k) = \hat{\boldsymbol{\phi}}(k|k-1) \\ + \Re\Big\{ \mathbf{G}(k)\Big[\mathbf{y}(k) - \hat{\boldsymbol{\Phi}}^{[r]}(k)\mathbf{H}\hat{\boldsymbol{\Phi}}^{[t]}(k)\bar{\mathbf{s}}(k) \Big] \Big\}, \end{aligned} \tag{30}$$

$$\begin{aligned} \mathbf{M}(k|k) = \Re\{ \mathbf{M}(k|k-1) \\ - \mathbf{G}(k)\mathbf{E}(\boldsymbol{\phi}(k|k-1))\mathbf{M}(k|k-1) \}, \end{aligned} \tag{31}$$

where

• $\mathbf{E}(\boldsymbol{\phi}(k|k-1)) = \mathbf{E}(\boldsymbol{\phi}(k))\big|_{\boldsymbol{\phi}(k)=\hat{\boldsymbol{\phi}}(k|k-1)}$,
• $\boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}(\boldsymbol{\phi}(k|k-1)) = \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}(k)}(\boldsymbol{\phi}(k))\big|_{\boldsymbol{\phi}(k)=\hat{\boldsymbol{\phi}}(k|k-1)}$,
• $\hat{\boldsymbol{\phi}}(k|k-1) = \Big[\hat{\phi}_1^{[t]}(k|k-1), \dots, \hat{\phi}_{N_t-1}^{[t]}(k|k-1), \hat{\phi}_1^{[r]}(k|k-1), \dots, \hat{\phi}_{N_r}^{[r]}(k|k-1) \Big]^T$ is the predicted state vector at the kth symbol,
• $\hat{\boldsymbol{\Phi}}^{[r]}(k) = \boldsymbol{\Phi}^{[r]}(k)\big|_{\boldsymbol{\phi}(k)=\hat{\boldsymbol{\phi}}(k|k-1)}$
  $\hat{\boldsymbol{\Phi}}^{[t]}(k) = \boldsymbol{\Phi}^{[t]}(k)\big|_{\boldsymbol{\phi}(k)=\hat{\boldsymbol{\phi}}(k|k-1)}$,
• $\mathbf{G}(k)$ is the $N \times N_r$ Kalman gain matrix,
• $\mathbf{M}(k|k)$ is the $N \times N$ EKF error covariance matrix, and
• $\bar{\mathbf{s}}(k)$ depends on $\rho_i(k)$, $\forall i$, in (17), which is evaluated at $\boldsymbol{\phi}(k) = \hat{\boldsymbol{\phi}}(k|k-1)$.

Before starting the EKF recursion (25)–(31), $\hat{\boldsymbol{\phi}}(0|0)$ and $\mathbf{M}(0|0)$ should be initialized with appropriate values. Here, the EKF is initialized with $\hat{\boldsymbol{\phi}}(0|0) = \mathbf{0}_{N\times 1}$ and $\mathbf{M}(0|0)$ is initialized with the covariance of the phase noise processes at time k, $\boldsymbol{\Sigma}$, defined in (10). This choice of initialization is justified since the phase noise estimates are assumed to be obtained via a training sequence prior to data transmission via the algorithm in [10].

The estimates $\hat{\boldsymbol{\phi}}(k)$ obtained via the proposed soft-input EKF can be refined using a soft-input EKS, which estimates $\hat{\boldsymbol{\phi}}(k|K)$, i.e., the estimate of $\boldsymbol{\phi}(k)$ given the observations and soft data symbol statistics of the whole frame [26]. In this approach, the proposed soft-input EKF is first applied symbol-by-symbol in the forward direction as shown in (25)–(31). Next, the proposed soft-input EKS refines the estimates, $\hat{\boldsymbol{\phi}}(k|K)$, via the following backward-recursive equations, for $k = K-1, K-2, \dots, 1$,

$$\hat{\boldsymbol{\phi}}(k|K) = \hat{\boldsymbol{\phi}}(k|k) + \mathbf{L}(k)\big(\hat{\boldsymbol{\phi}}(k+1|K) - \hat{\boldsymbol{\phi}}(k+1|k)\big), \tag{32}$$

$$\begin{aligned} \bar{\mathbf{M}}(k|K) = \mathbf{M}(k|k) \\ + \mathbf{L}(k)\big(\bar{\mathbf{M}}(k+1|K) - \mathbf{M}(k+1|k)\big)\mathbf{L}^H(k), \end{aligned} \tag{33}$$

where $\mathbf{L}(k) \triangleq \mathbf{M}(k|k)\big(\mathbf{M}(k+1|k)\big)^{-1}$ and $\bar{\mathbf{M}}(k|K)$ is the EKS error covariance matrix.

C. Complexity Analysis

In this paper, computational complexity is defined as the number of complex *additions plus multiplications* required to obtain the phase noise estimates, $\hat{\boldsymbol{\phi}}(k)$, in the kth symbol interval[6]. Throughout this subsection the superscripts $(\cdot)^{[M]}$ and $(\cdot)^{[A]}$ are used to denote the number of multiplications and additions required by each algorithm, respectively. The computational complexity of the online MAP algorithm, denoted by $C_{\text{On-MAP}} = C_{\text{On-MAP}}^{[M]} + C_{\text{On-MAP}}^{[A]}$, can be determined as

$$\begin{aligned} & C_{\text{On-MAP}}^{[M]} \\ & = \mathcal{N}N\frac{2\pi}{\kappa}\Big\{ \underbrace{N_r(N_tN + 2N_t+1)+1}_{\text{first factor in (23)}} + \underbrace{2N^2+1}_{\text{second factor in (23)}} \Big\}, \end{aligned} \tag{34a}$$

$$\begin{aligned} & C_{\text{On-MAP}}^{[A]} \\ & = \mathcal{N}N\frac{2\pi}{\kappa}\Big\{ \underbrace{N_r(N_tN + 1)-1}_{\text{first factor in (23)}} + \underbrace{2N(N-1)}_{\text{second factor in (23)}} \\ & \qquad\qquad + \underbrace{1}_{\text{summation of 2 factors}} \Big\}, \end{aligned} \tag{34b}$$

where $\mathcal{N}$ denotes the number of alternating projection cycles used, and κ denotes the step size used for the exhaustive search in (23). Similarly, the computational complexity of the proposed offline MAP algorithm, denoted by $C_{\text{Off-MAP}} = C_{\text{Off-MAP}}^{[M]} + C_{\text{Off-MAP}}^{[A]}$, can be also calculated as

$$\begin{aligned} & C_{\text{Off-MAP}}^{[M]} \\ & = \frac{1}{K}\mathcal{N}KN\frac{2\pi}{\kappa}K\Big\{ \underbrace{N_r(N_tN + 2N_t + 1) + 1}_{\text{first factor in (22)}} \\ & \qquad\qquad + \underbrace{2N^2+1}_{\text{second factor in (22)}} \Big\}, \end{aligned} \tag{35a}$$

---

[6]In this paper, finding the inverse of an $N \times N$ matrix is assumed to require N^3 arithmetic operations [59].

$$C_{\text{Off-MAP}}^{[A]}$$

$$= \frac{1}{K}\mathcal{N}KN\frac{2\pi}{\kappa}\left\{K\left(\underbrace{N_r(N_tN+1)-1}_{\text{first factor in (22)}}+\underbrace{2N(N-1)}_{\text{second factor in (22)}}\right.\right.$$

$$\left.\left.+\underbrace{1}_{\text{summation of 2 factors}}\right)+(K-1)\right\}. \quad (35b)$$

Furthermore, the complexity of the proposed soft-input EKF algorithm, $C_{\text{EKF}} = C_{\text{EKF}}^{[M]} + C_{\text{EKF}}^{[A]}$, can be written as

$$C_{\text{EKF}}^{[M]} = \underbrace{2N_rN(N+N_r)+N_r^3}_{(29)}+\underbrace{N_r(N+N_r)}_{(30)}$$

$$+\underbrace{N^2(N_r+N)}_{(31)}+\underbrace{4N-2N_r}_{(25)}+\underbrace{2N_rN_t^2+N_t}_{(20)}$$

$$+\underbrace{S(N_rN_t+N+4)+N_rN_t(N+1)}_{\bar{s}(k)\text{ in }(30)}, \quad (36a)$$

$$C_{\text{EKF}}^{[A]} = \underbrace{N^2}_{(28)}+\underbrace{NN_r(2N+2N_r-3)+N_r^3}_{(29)}$$

$$+\underbrace{N_r(N+N_t)}_{(30)}+\underbrace{N^2(N_r+N-1)}_{(31)}$$

$$+\underbrace{N_r(N_t-1)}_{(25)}+\underbrace{2N_rN_t(N_t-1)+N_r+N_t}_{(20)}$$

$$+\underbrace{S(N_rN_t+N+1)+N_rN_t(N-1)-(N_t+1)}_{\bar{s}(k)\text{ in }(30)}. \quad (36b)$$

Finally, the computational complexity of the proposed soft-input EKF-EKS algorithm, $C_{\text{EKF-EKS}} = C_{\text{EKF-EKS}}^{[M]} + C_{\text{EKF-EKS}}^{[A]}$, can be written as

$$C_{\text{EKF-EKS}}^{[M]} = C_{\text{EKF}}^{[M]}+\underbrace{N^2}_{(32)}+\underbrace{2N^3}_{\mathbf{L}(k)\text{ in }(32)}, \quad (37a)$$

$$C_{\text{EKF-EKS}}^{[A]} = C_{\text{EKF}}^{[A]}+\underbrace{N(1+N)}_{(32)}+\underbrace{N^2(2N-1)}_{\mathbf{L}(k)\text{ in }(32)}. \quad (37b)$$

Remark 3: In Table I, the computational complexity of the proposed online MAP, offline MAP, soft-input EKF, soft-input EKF-EKS, and the algorithms in [10] and [15], denoted by $C_{\text{On-MAP}}$, $C_{\text{Off-MAP}}$, C_{EKF}, $C_{\text{EKF-EKS}}$, $C_{[10]}$, and $C_{[15]}$, respectively, for 2×2, 4×4, and 8×8 MIMO systems are compared against one another. In this comparison, $C_{[10]}$ is evaluated via [10, Eq. (41)], $C_{[15]}$ is evaluated via [15, Eq. (32)], and $C_{\text{On-MAP}}$, $C_{\text{Off-MAP}}$, C_{EKF}, and $C_{\text{EKF-EKS}}$ are evaluated by (34), (35), (36) and (37), respectively. As indicated in Section V, to ensure that the MAP estimator's MSE is close to the CRLB, we set the step size $\kappa = 10^{-3}$ and $\mathcal{N} = 4$ in (34a), and $K = 20$ in (35). Table I shows that the proposed soft-input EKF and EKF-EKS are computationally more efficient than the algorithms in [10], [15], and the proposed online and offline MAP estimators. For example, for a 4×4 MIMO system, the proposed soft-input EKF-EKS algorithm

TABLE I
COMPUTATIONAL COMPLEXITY OF MAP, EKF, EKF-EKS, AND THE
ALGORITHMS IN [10] AND [15] FOR DIFFERENT NUMBERS OF ANTENNAS

MIMO	$C_{\text{On-MAP}}$	$C_{\text{Off-MAP}}$	C_{EKF}	$C_{\text{EKF-EKS}}$	$C_{[10]}$	$C_{[15]}$
2×2	5.1e6	1.0e8	3.6e2	4.8e2	1.2e3	1.7e4
4×4	7.8e7	1.5e9	3.7e3	5.1e3	5.6e5	5.4e5
8×8	1.1e9	2.2e10	6.8e4	8.1e4	2.8e8	1.7e7

is 3.0×10^5, 1.2×10^2, and 1.0×10^2 times less complex than the proposed offline MAP estimator in (22), the algorithm in [10], and the scheme in [15], respectively. This advantage with respect to the MAP estimator is expected since the MAP estimator requires an expensive exhaustive search to obtain the phase noise estimates. In addition, the proposed EKF-EKS approach is less complex than the phase tracking algorithm in [10], [15], since it tracks fewer phase noise processes at every time instant k.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of the proposed estimators and compare it with the BCRLB. Subsequently, the BER of a MIMO system employing the proposed phase noise tracking scheme is investigated. Here, the MIMO channel gains, $\mathbf{H}$, and phase noise parameters at the beginning of each frame are obtained using the algorithm in [10], where a total of 2 and 4 symbols are used for training for 2×2 and 4×4 MIMO systems, respectively. Throughout this section it is assumed that $\sigma_\delta^2 = \sigma_{\delta_m^{[t]}}^2 = \sigma_{\delta_n^{[r]}}^2 \ \forall \ m, n$, and $\sigma_w^2 = 1/\text{SNR}$. The MIMO channel matrix is generated as a sum of *line-of-sight (LoS)* and *non-line-of-sight (NLoS)* components such that the overall channel matrix, $\mathbf{H}$, is given by [17]

$$\mathbf{H} = \sqrt{\frac{\xi}{\xi+1}}\mathbf{H}_{\text{LoS}} + \sqrt{\frac{1}{\xi+1}}\mathbf{H}_{\text{NLoS}}, \quad (38)$$

where ξ denotes the Rician factor [60, p.52].[7] The elements of $\mathbf{H}_{\text{LoS}}$ are generated according to the model in [17] and the elements of $\mathbf{H}_{\text{NLoS}}$ are modeled as independent and identically distributed complex Gaussian random variables with $\mathcal{CN}(0,1)$. Given that the estimation ranges of the proposed soft-input MAP and EKF estimators are limited to $[-\pi, \pi)$, the phase unwrapping algorithm in [18] is applied here, where phase noise estimates for prior symbols are used in combination with the phase noise variance to unwrap the estimate for the current symbol. To evaluate the average BER performance, a minimum of 10^5 Monte-Carlo trials are used, where 200 symbols in a frame are transmitted from each antenna per trial. The soft-input estimation and decoding are carried out according to *Steps 1–3* in *Remark 1*. The MSE and BER performances of the proposed estimators and the overall MIMO system, respectively, are presented in Sections V-A and V-B.

A. Estimation Performance

Unless otherwise specified, only in this subsection, a channel realization, $\mathbf{H} = [0.9928 + 0.2920j, -0.6541 - 1.2625j; 1.2740 - 0.2759j, 0.3207 - 2.0030j]$, drawn from the model in (38) for a 2×2 MIMO system, is used for all

---

[7]Note that $\xi = -\infty$ dB and $\xi = \infty$ dB result in pure NLoS and LoS of channels, respectively [17].
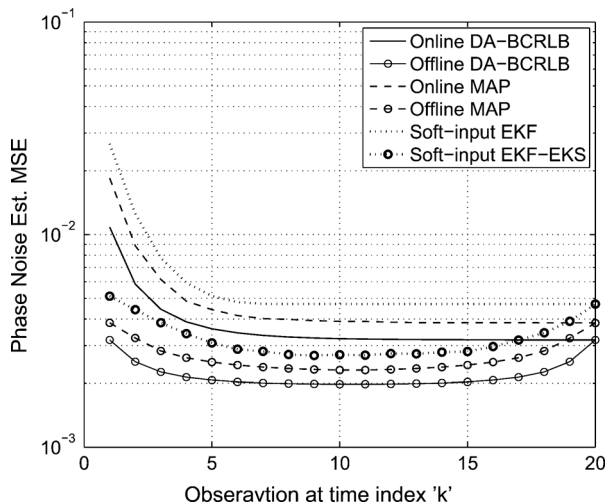
Fig. 4. Phase noise estimation MSEs and BCRLB versus observation index, k, for a 2×2 MIMO system with $\sigma_\delta^2 = 10^{-3} \ \mathrm{rad}^2$, SNR $= 10$ dB, and $K = 20$.

Fig. 5. Online MAP estimator's MSE and DA-BCRLB versus SNR for a 2×2 MIMO system with $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \ \mathrm{rad}^2$ and $K = 20$.

Fig. 6. EKF's MSEs and the DA-BCRLB versus SNR for a 2×2 MIMO system with $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \ \mathrm{rad}^2$ and $K = 20$.

simulations as in [56], [61], while the AWGN and phase noise values change independently from symbol to symbol. BPSK modulation is used for the data symbols. The step size $\kappa = 10^{-3}$ and $\mathcal{N} = 4$ are selected to ensure that the MSE of the proposed soft-input online and offline MAP estimators is close to the DA-BCRLB. The remaining sets of parameters for this subsection are selected to be the same as that for Fig. 2. Only the MSE for the estimation of $\phi_{N_r}^{[r]}(k)$, $\forall k$ is presented, since similar results are obtained for the remaining phase noise processes.

Fig. 4 plots the DA-BCRLBs and the *offline* and *online* phase noise estimation MSEs for the proposed estimators versus time index k, for $k = [1, \ldots, 20]$, SNR $= 10$ dB, and $\sigma_\delta^2 = 10^{-3} \ \mathrm{rad}^2$. The results in Fig. 4 indicate that the performances of the proposed online and offline MAP estimators are close to their respective BCRLBs while the MSEs of the proposed soft-input EKF and EKF-EKS estimators are slightly higher than their MAP counterparts. As indicated in more detail in Fig. 6 below, this result is expected given that unlike the soft-input MAP estimators, the proposed soft-input EKF and EKF-EKS algorithms both apply the phase noise estimates at the previous time instant and a Taylor series approximation to linearize the observation equations. In addition, Fig. 4 illustrates that the proposed offline MAP and EKF-EKS estimators' MSEs are lower than the online DA-BCRLB. This result is justifiable, given that both estimators in offline mode take advantage of past and future observation symbols over the entire frame to estimate highly correlated phase noise values. Therefore, the online BCRLB and the performance of the online MAP and EKF estimators exhibit higher MSE values when compared to their offline counterparts. In light of this finding, it is important to note that as shown in Table I the proposed online estimators have lower computational complexities and delay than the offline estimators.

Fig. 5 plots the MSE performances of the proposed MAP estimators against the DA-BCRLBs as functions of the SNR for phase noise variances $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \ \mathrm{rad}^2$. Fig. 5 illustrates that at low SNR, the BCRLB is more dependent on
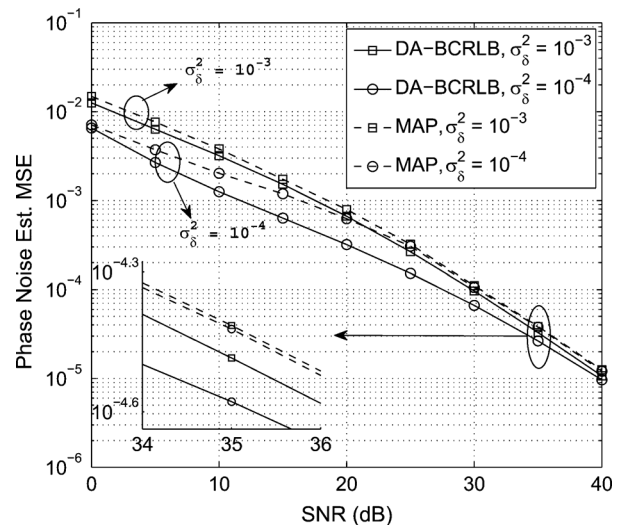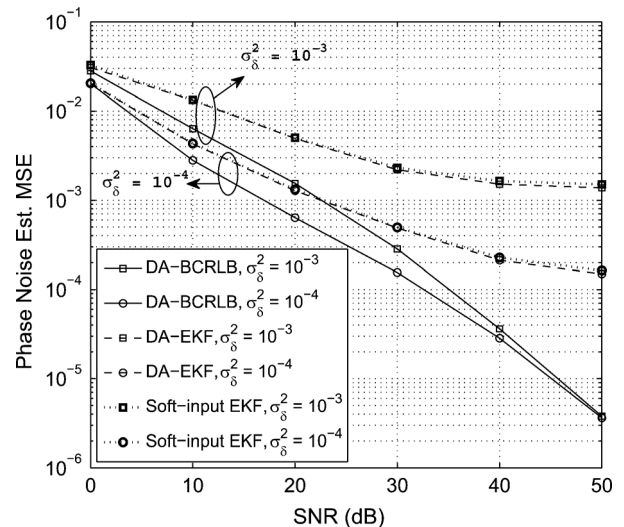
the variance of the phase noise process, where it is lower for a lower phase noise variance, e.g., $\sigma_\delta^2 = 10^{-4} \ \mathrm{rad}^2$ compared to $\sigma_\delta^2 = 10^{-3} \ \mathrm{rad}^2$. However, at high SNR, the two curves approach each other. This can be explained by the inherent structure of the DA-BCRLBs in (12) and (14), where at low SNR, they are more dependent on the a priori distribution of $\phi(k)$, i.e., the state model. As a result, a lower phase noise variance results in lower BCRLBs, whereas at high SNR, the observations are trusted more and the BCRLBs are dominated by the variance of the AWGN process. It should be further noted that the MSEs of the proposed soft-input MAP estimators are close to the derived DA-BCRLBs at medium-to-high SNRs, for different phase noise variances.

Fig. 6 plots the MSE performance of the proposed soft-input EKF estimator and the DA-BCRLB as a function of SNR for phase noise variances $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \ \mathrm{rad}^2$ at time instant $k = K = 20$. In this case, independent Rayleigh fading channels, i.e., $\xi = -\infty$ dB, are generated for each frame. Note

that as discussed above and illustrated in Fig. 4, the EKF and EKF-EKS algorithms' MSEs overlap one another when estimating the phase noise process corresponding to the last symbol in a frame, i.e., for $k = K$. In order to find an even tighter lower bound on the performance of the proposed soft-input EKF, its MSE in the data-aided mode is also plotted in this figure.[8] It can be seen from Fig. 6 that the MSE of the proposed soft-input EKF is close to that of the DA-EKF for the whole range of SNR. In addition, this figure indicates that the MSE of this estimator exhibits an error floor at high SNR, which is directly related to the variance of the phase noise process. This error floor is caused by the soft-input EKF algorithm's structure in (30), where the estimates of the phase noise processes at the kth symbol interval depend on the phase noise estimates obtained at the $(k-1)$th time instant. Therefore, even when operating in the data-aided mode and in the absence of AWGN, i.e., $\mathrm{SNR} = \infty$, the estimation accuracy of the proposed EKF estimator is limited by the variance of the phase noise innovations. This error floor can be also attributed to the fact that the EKF is not an optimal minimum mean square filter due to the application of a first order Taylor series approximation to linearize the observation model [46].

B. MIMO System Performance

In this section, the BER of an uncoded MIMO system versus SNR per symbol for different phase noise variances, different numbers of antennas, and various modulation schemes is investigated and is compared against that of [10] and [15]. Moreover, the BER of a MIMO system with *perfect* knowledge of channels and phase noise values is also evaluated and used as a benchmark. The received signals are decoded using the soft-decoding approach described in *Remark 1*. The frame length is set to $K = 200$ symbols. Since the scheme in [15] requires transmission of orthogonal pilot symbols to obtain its BER plots, 20 and 40 pilots are transmitted for 2×2 and 4×4 MIMO systems, respectively. Furthermore, in [15], using orthogonal pilot symbols, the sum phase noise processes, i.e., $\theta_{m,n}(k) = \theta_m(k) + \theta_n(k)$, are estimated. Different channels are generated for each frame. Unless otherwise specified, ξ is set to $-\infty$ dB in this section, which corresponds to Rayleigh fading channels. Finally, the BER for the proposed MAP estimator is not investigated due to its high complexity as shown in Section IV-C.

Fig. 7 plots the BER of a 2×2 MIMO system using BPSK modulation for phase noise variances $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \; \mathrm{rad}^2$. This figure illustrates that when using the proposed soft-input EKF-EKS estimator the BER of a MIMO system is only 3 dB away from the ideal case of perfect channel and phase noise knowledge. Note that this benchmark plot is independent of the phase noise variance since it is based on the exact knowledge of the phase noise values. The results in Fig. 7 also indicates that at high SNR the overall system suffers from an error floor[9], which is due to the limitation of the proposed soft-input EKF-EKS algorithm, see Fig. 6. However, when compared to the algorithms in [10], [15], the proposed receiver's error floor appears at

[8]For the DA-EKF, $\boldsymbol{\Sigma}_{\bar{\mathbf{w}}(k)}\left(\boldsymbol{\phi}(k|k-1)\right) = \sigma_w^2 \mathbf{I}_{N_r \times N_r}$ in (29) and $\bar{\mathbf{s}}(k) = \mathbf{s}(k)$ in (30).

[9]In Fig. 7, the error floor for the BER plot for phase noise variance of $\sigma_\delta^2 = 10^{-4} \; \mathrm{rad}^2$, appears at higher SNRs, e.g., $\mathrm{SNR} > 40$ dB.
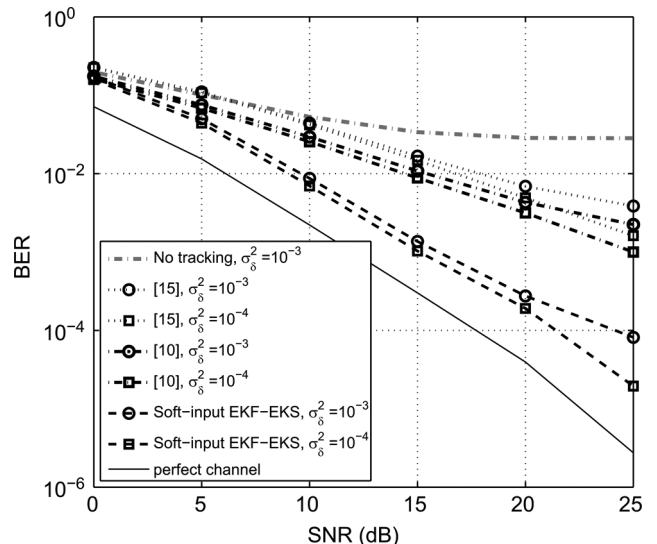

Fig. 7. BER of a 2×2 MIMO system for phase noise variances, $\sigma_\delta^2 = [10^{-3}, 10^{-4}] \; \mathrm{rad}^2$, Rayleigh fading and BPSK modulation.
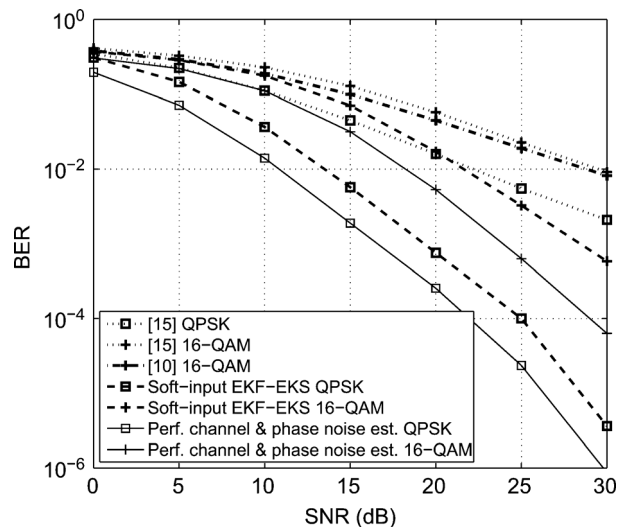

Fig. 8. BER of a 2×2 MIMO system for QPSK and 16QAM modulations for Rayleigh fading and $\sigma_\delta^2 = 10^{-4} \; \mathrm{rad}^2$.

considerably lower BER values. For example, to achieve a BER of 1.4×10^{-3} for a phase noise variance of $\sigma_\delta^2 = 10^{-3} \; \mathrm{rad}^2$, the soft-input EKF-EKS achieves an 18 dB performance gain compared to [10], [15]. This gain is achieved while reducing synchronization overhead.

Fig. 8 depicts the BER performance of a 2×2 MIMO system for higher order modulations, e.g., *quadrature phase-shift keying* (QPSK) and 16-QAM, for phase noise variance $\sigma_\delta^2 = 10^{-4} \; \mathrm{rad}^2$. This figure shows that the BER of a MIMO system employing the proposed soft-input EKF-EKS estimator is close to the benchmark curve, with a performance gap of approximately 2 dB for QPSK and 16-QAM at moderate SNR values. Moreover, the results in Fig. 8 indicate that compared to the algorithm in [15], the proposed soft-input EKF-EKS receiver can support higher order modulations, which can enhance the capacity and throughput of MIMO systems to a great extent.

As the number of antennas in a MIMO system increases, a larger number of phase noise processes need to be tracked by
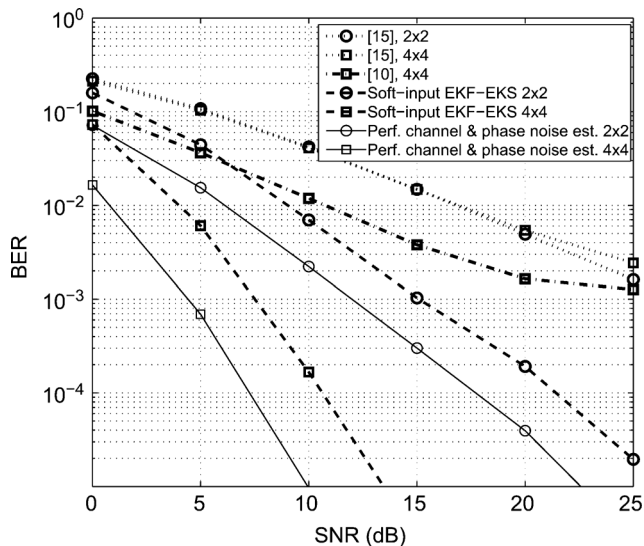
Fig. 9. BER of a MIMO system for different N_t and N_r values, for BPSK modulation, Rayleigh fading, and $\sigma_\delta^2 = 10^{-4}\,\mathrm{rad}^2$.
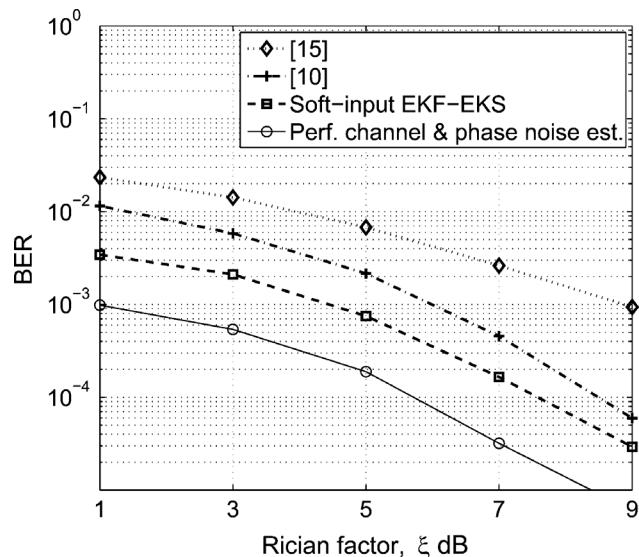

Fig. 10. BER of a 2×2 MIMO system as a function of the Rician factor, ξ, for BPSK modulation, $\sigma_\delta^2 = 10^{-4}\,\mathrm{rad}^2$, and $\mathrm{SNR} = 10$ dB.

the proposed EKF-EKS estimator, which in turn can negatively affect its performance. Thus, in Fig. 9, we compare the BER performances of 2×2 and 4×4 MIMO systems in the presence of phase noise. In this setup, BPSK modulation is applied and the phase noise variance is set to $\sigma_\delta^2 = 10^{-4}\,\mathrm{rad}^2$. Fig. 9 shows that compared to the ideal scenario, the BER of the proposed soft-input EKF-EKS receiver is only 3 and 4 dB higher in the low-to-moderate SNR region for 2×2 and 4×4 MIMO systems, respectively. Moreover, the proposed soft-input EKF-EKS estimator outperforms the phase tracking approach in [15] and [10] by large margins.

Fig. 10 illustrates the BER of a 2×2 MIMO system as a function of Rician factor, ξ, for $\mathrm{SNR} = 10$ dB. Here, BPSK modulation is employed and the phase noise variance is set to $\sigma_\delta^2 = 10^{-4}\,\mathrm{rad}^2$. As anticipated, this figure indicates that the BER of a MIMO system decreases as the Rician factor increases due to a stronger LoS or deterministic channel compo-

nent. Fig. 10 also shows that the BER of a MIMO system employing the proposed soft-input EKF-EKS receiver is just two times higher than the benchmark BER and that performance gap remains constant for different values of Rician factor. However, the performance gap between the BER for the algorithm in [15] and the benchmark scheme increases with increasing Rician factor. Fig. 10 also demonstrates that for a Rician factor of $\xi = 7$, the BER for the proposed soft-input EKF-EKS receiver is approximately 15.7 and 2.8 times lower than that of the algorithms in [15] and [10], respectively.

VI. Conclusions

In this paper a new signal model for the estimation of the phase noise in MIMO systems is outlined and new exact expressions for the online and offline DA- and NDA-BCRLBs and soft-input MAP estimators for the estimation of phase noise are derived. Numerical results demonstrate that the proposed MAP estimators' MSE performances are close to the derived BCRLBs at moderate-to-high SNRs. Next, a novel soft-input EKF-EKS based estimator is proposed. Extensive simulations indicate that over a wide range of SNR values the BER performance of a MIMO system using the proposed soft-input EKF-EKS estimator is only slightly higher than the BER of the idealistic scenario, i.e., when assuming perfect channel and phase noise knowledge at the receiver. Moreover, results show that the proposed soft-input EKF-EKS estimator enables application of higher order modulations and larger numbers of antennas, which can in turn improve the bandwidth efficiency of wireless systems.

Although the proposed scheme cannot be directly applied to frequency selective channels, the principles and methodologies proposed here can be used to develop multiple phase noise estimation algorithms for such channels, e.g., using OFDM and *orthogonal frequency division multiple access (OFDMA)*. Similar to single carrier systems, OFDM and OFDMA systems may be affected by multiple multiplicative phase noise processes [36], i.e., common phase errors. Thus, the proposed algorithms can be applied to such systems.[10]

Appendix A
Derivation of BIM

In this appendix, a closed-form expression for the BIM, $\mathbf{B}$ in (6), is derived.

A. Derivation of $\mathbb{E}_\phi[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})]$

Given the signal model in (2) and the assumptions in Section II, the log-likelihood function, $\log p(\mathbf{y}|\boldsymbol{\phi},\mathbf{s})$, is

$$\log p(\mathbf{y}|\boldsymbol{\phi},\mathbf{s}) = \sum_{k=1}^{K} \log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))$$

$$= \sum_{k=1}^{K} -N_r \log(\pi\sigma_w^2)$$

$$- \frac{1}{\sigma_w^2}\left\|\mathbf{y}(k) - \boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{s}(k)\right\|^2. \quad (A.1)$$

[10]Addressing this specific problem is beyond the scope of this paper and is the subject of future work.

To determine $\mathbb{E}_{\boldsymbol{\phi}}[\mathbf{F}(\boldsymbol{\phi},\mathbf{s})] = \mathbb{E}_{\mathbf{y},\boldsymbol{\phi}|\mathbf{s}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}}\log p(\mathbf{y}|\boldsymbol{\phi},\mathbf{s})\right]$, a closed-form expression for $\mathbb{E}_{\mathbf{y},\boldsymbol{\phi}|\mathbf{s}}\left[-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)}\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))\right]$ is obtained for $k = 1,\ldots,K$. Using (A.1), $\boldsymbol{F} \triangleq -\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))$ can be expressed as

$$\boldsymbol{F} = N_r \log(\pi\sigma_w^2) + \frac{\mathbf{y}^H(k)\mathbf{y}(k)}{\sigma_w^2}$$
$$+ \frac{1}{\sigma_w^2}\Big(-2\Re\{\mathbf{y}^H(k)\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{s}(k)\}$$
$$+ \mathbf{s}^H(k)\big(\boldsymbol{\Phi}^{[t]}(k)\big)^H\mathbf{H}^H\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{s}(k)\Big). \quad \text{(A.2)}$$

In order to determine the $N \times N$ Hessian matrix, $-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)}\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))$, the derivatives, $\frac{\partial \boldsymbol{F}}{\partial \phi_m^{[t]}(k)}$, $\frac{\partial \boldsymbol{F}}{\partial \phi_n^{[r]}(k)}$, $\frac{\partial^2 \boldsymbol{F}}{\left(\partial \phi_m^{[t]}(k)\right)^2}$, $\frac{\partial^2 \boldsymbol{F}}{\left(\partial \phi_n^{[r]}(k)\right)^2}$, $\frac{\partial^2 \boldsymbol{F}}{\partial \phi_n^{[r]}(k)\partial \phi_m^{[t]}(k)}$, $\frac{\partial^2 \boldsymbol{F}}{\partial \phi_{\tilde{m}}^{[t]}(k)\partial \phi_m^{[t]}(k)}$, and $\frac{\partial^2 \boldsymbol{F}}{\partial \phi_{\tilde{n}}^{[r]}(k)\partial \phi_n^{[r]}(k)}$, for $m = 1,\ldots,N_t-1$, $\tilde{m} = 1,\ldots,N_t-1$, $n = 1,\ldots,N_r$, and $\tilde{n} = 1,\ldots,N_r$, are calculated as

$$\frac{\partial \boldsymbol{F}}{\partial \phi_m^{[t]}(k)} = \frac{1}{\sigma_w^2}\Big(-2\Re\{\mathbf{y}^H(k)\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\mathbf{s}(k)$$
$$- \mathbf{s}^H(k)\big(\boldsymbol{\Phi}^{[t]}(k)\big)^H\mathbf{H}^H\mathbf{H}\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\mathbf{s}(k)\}\Big), \tag{A.3}$$

$$\frac{\partial \boldsymbol{F}}{\partial \phi_n^{[r]}(k)} = \frac{1}{\sigma_w^2}\Big(-2\Re\{\mathbf{y}^H(k)\dot{\boldsymbol{\Phi}}_n^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{s}(k)\}\Big), \tag{A.4}$$

$$\frac{\partial^2 \boldsymbol{F}}{\left(\partial \phi_m^{[t]}(k)\right)^2} = \frac{1}{\sigma_w^2}\Big(-2\Re\{\mathbf{y}^H(k)\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\ddot{\boldsymbol{\Phi}}_m^{[t]}(k)\mathbf{s}(k)$$
$$- \mathbf{s}^H(k)\big(\boldsymbol{\Phi}^{[t]}(k)\big)^H\mathbf{H}^H\mathbf{H}\ddot{\boldsymbol{\Phi}}_m^{[t]}(k)\mathbf{s}(k)$$
$$- \mathbf{s}^H(k)\big(\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\big)^H\mathbf{H}^H\mathbf{H}\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\mathbf{s}(k)\}\Big), \tag{A.5}$$

$$\frac{\partial^2 \boldsymbol{F}}{\left(\partial \phi_n^{[r]}(k)\right)^2} = \frac{1}{\sigma_w^2}\Big(-2\Re\{\mathbf{y}^H(k)\ddot{\boldsymbol{\Phi}}_n^{[r]}(k)\mathbf{H}\boldsymbol{\Phi}^{[t]}(k)\mathbf{s}(k)\}\Big), \tag{A.6}$$

$$\frac{\partial^2 \boldsymbol{F}}{\partial \phi_n^{[r]}(k)\partial \phi_m^{[t]}(k)} = \frac{\partial^2 \boldsymbol{F}}{\partial \phi_m^{[t]}(k)\partial \phi_n^{[r]}(k)}$$
$$= \frac{1}{\sigma_w^2}\Big(-2\Re\{\mathbf{y}^H(k)\dot{\boldsymbol{\Phi}}_n^{[r]}(k) \tag{A.7}$$
$$\times \mathbf{H}\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\mathbf{s}(k)\}\Big),$$

$$\frac{\partial^2 \boldsymbol{F}}{\partial \phi_{\tilde{m}}^{[t]}(k)\partial \phi_m^{[t]}(k)} = \frac{1}{\sigma_w^2}\Big(2\Re\{\mathbf{s}^H(k)\big(\dot{\boldsymbol{\Phi}}_{\tilde{m}}^{[t]}(k)\big)^H\mathbf{H}^H$$
$$\times \mathbf{H}\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\mathbf{s}(k)\}\Big), \quad \tilde{m} \neq m, \tag{A.8}$$

$$\frac{\partial^2 \boldsymbol{F}}{\partial \phi_{\tilde{n}}^{[r]}(k)\partial \phi_n^{[r]}(k)} = 0, \quad \tilde{n} \neq n \tag{A.9}$$

where

- $\dot{\boldsymbol{\Phi}}_m^{[t]}(k) = \frac{\partial \boldsymbol{\Phi}^{[t]}(k)}{\partial \phi_m^{[t]}(k)}$ is an $N_t \times N_t$ matrix with $\left[\dot{\boldsymbol{\Phi}}_m^{[t]}(k)\right]_{m,m} = je^{j\phi_m^{[t]}(k)}$, $\forall m$, and zeros elsewhere,

- $\ddot{\boldsymbol{\Phi}}_m^{[t]}(k) = \frac{\partial^2 \boldsymbol{\Phi}^{[t]}(k)}{\partial\left(\phi_m^{[t]}(k)\right)^2}$ is an $N_t \times N_t$ matrix with $\left[\ddot{\boldsymbol{\Phi}}_m^{[t]}(k)\right]_{m,m} = -e^{j\phi_m^{[t]}(k)}$, $\forall m$, and zeros elsewhere,

- $\dot{\boldsymbol{\Phi}}_n^{[r]}(k) = \frac{\partial \boldsymbol{\Phi}^{[r]}(k)}{\partial \phi_n^{[r]}(k)}$ is an $N_r \times N_r$ matrix with $\left[\dot{\boldsymbol{\Phi}}_n^{[r]}(k)\right]_{n,n} = je^{j\phi_n^{[r]}(k)}$, $\forall n$, and zeros elsewhere else, and

- $\ddot{\boldsymbol{\Phi}}_n^{[r]}(k) = \frac{\partial^2 \boldsymbol{\Phi}^{[r]}(k)}{\partial\left(\phi_n^{[r]}(k)\right)^2}$ is an $N_r \times N_r$ matrix with $\left[\ddot{\boldsymbol{\Phi}}_n^{[r]}(k)\right]_{n,n} = -e^{j\phi_n^{[r]}(k)}$, $\forall n$, and zeros elsewhere.

In order to evaluate the expectation of the Hessian matrix, $-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)}\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))$, over $\bar{\mathbf{y}}$ and $\boldsymbol{\phi}$,

$$\boldsymbol{\Pi}(k) \triangleq \mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\phi}|\mathbf{s}}\left[-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)}\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))\right],$$

the PDF of $\boldsymbol{\phi}(k)$, which depends on the distribution of $\boldsymbol{\theta}(k)$ needs to be determined. Based on the assumptions in Section II, in (1), $\theta_n^{[r]}(k)$ and $\theta_m^{[t]}(k)$ at any time instant k, for $n = 1,\ldots,N_r$ and $m = 1,\ldots,N_t$, are uniformly independently and identically distributed over the range $(-\pi,\pi)$[11]. In addition, the PDFs of $\phi_m^{[t]}(k) \triangleq \theta_m^{[t]}(k) - \theta_{N_t}^{[t]}(k)$ and $\phi_n^{[r]}(k) \triangleq \theta_n^{[r]}(k) + \theta_{N_t}^{[t]}(k)$ are given by the convolution of the PDFs of $\left\{\theta_m^{[t]}(k),\theta_{N_t}^{[t]}(k)\right\}$ and $\left\{\theta_n^{[r]}(k),\theta_{N_t}^{[t]}(k)\right\}$, respectively,[12] which are straightforwardly determined to be triangular PDFs in the range of $(-2\pi,2\pi)$. Moreover, given the signal model in (2), the PDFs of $\phi_m^{[t]}(k)$ and $\phi_n^{[r]}(k)$, $\forall m, n$, can be equivalently wrapped, resulting in uniform PDFs over $(-\pi,\pi)$ for $\phi_m^{[t]}(k)$ and $\phi_n^{[r]}(k)$, $\forall m, n$.

Using (2), (A.3)–(A.9), and the statistics of $\mathbf{w}(k)$ and $\boldsymbol{\phi}(k)$, the expectation of the ℓth row and $\bar{\ell}$th column element of the Hessian matrix $-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)}\log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))$ with respect to $\mathbf{y}$ and $\boldsymbol{\phi}$, $[\boldsymbol{\Pi}(k)]_{\ell,\bar{\ell}}$, can be determined as

$$[\boldsymbol{\Pi}(k)]_{\ell,\bar{\ell}} = \frac{2}{\sigma_w^2}\mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\phi}|\mathbf{s}}\Big[-\Re\{\mathbf{w}^H(k)\boldsymbol{\Phi}^{[r]}(k)\mathbf{H}\ddot{\boldsymbol{\Phi}}_\ell^{[t]}(k)\mathbf{s}(k)$$
$$- \mathbf{s}^H(k)\big(\dot{\boldsymbol{\Phi}}_\ell^{[t]}(k)\big)^H\mathbf{H}^H\mathbf{H}\dot{\boldsymbol{\Phi}}_\ell^{[t]}(k)\mathbf{s}(k)\}\Big]$$
$$= \frac{2}{\sigma_w^2}\mathbb{E}_{\boldsymbol{\phi}|\mathbf{s}}\Big[\Re\{-js_\ell^*(k)e^{-j\phi_\ell^{[t]}(k)}([\mathbf{H}]_{1:N_r,\ell})^H$$
$$\times ([\mathbf{H}]_{1:N_r,\ell})je^{j\phi_\ell^{[t]}(k)}s_\ell(k)\}\Big]$$
$$= \frac{2}{\sigma_w^2}|s_\ell(k)|^2\sum_{n=1}^{N_r}|h_{n,l}|^2, \quad \ell=\bar{\ell}=1,\ldots,N_t-1 \tag{A.10}$$

$$[\boldsymbol{\Pi}(k)]_{\ell,\bar{\ell}} = \frac{1}{\sigma_w^2}\mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\phi}|\mathbf{s}}\Big[2\Re\{\mathbf{s}^H(k)\big(\dot{\boldsymbol{\Phi}}_{\bar{\ell}}^{[t]}(k)\big)^H\mathbf{H}^H\mathbf{H}\dot{\boldsymbol{\Phi}}_\ell^{[t]}(k)\mathbf{s}(k)\}\Big]$$
$$= \frac{1}{\sigma_w^2}\mathbb{E}_{\boldsymbol{\phi}|\mathbf{s}}\Big[2\Re\{-js_\ell^*(k)e^{-j\phi_{\bar{\ell}}^{[t]}(k)}([\mathbf{H}]_{1:N_r,\ell})^H$$
$$\times ([\mathbf{H}]_{1:N_r,\ell})je^{j\phi_\ell^{[t]}(k)}s_\ell(k)\}\Big] = 0,$$
$$\ell = 1,\ldots,N_t-1, \quad \bar{\ell} = 1,\ldots,N_t-1, \quad \ell \neq \bar{\ell} \tag{A.11}$$

[11]Assuming no prior information about phase noise values at time instant $k-1$, $\theta_n^{[r]}(k)$ and $\theta_{\bar{m}}^{[t]}(k)$, $\forall \bar{m}, n, k$, can assume any value between $-\pi$ and π with equal probability.

[12]The PDF of the sum of two statistically independent random variables is given by the convolution of their individual PDFs [62].

$$[\mathbf{\Pi}(k)]_{\ell,\bar{\ell}} = \frac{2}{\sigma_w^2} \mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\phi}|\mathbf{s}} \left[-\Re\left\{ \mathbf{y}^H(k)\ddot{\mathbf{\Phi}}_i^{[r]}(k)\mathbf{H}\mathbf{\Phi}^{[t]}(k)\mathbf{s}(k) \right\} \right]$$

$$= \frac{2}{\sigma_w^2} \mathbb{E}_{\boldsymbol{\phi}|\mathbf{s}} \left[-\Re\left\{ \mathbf{s}^H(k)\left(\mathbf{\Phi}^{[t]}(k)\right)^H \mathbf{H}^H \left(\mathbf{\Phi}^{[r]}(k)\right)^H \right.\right.$$
$$\left.\left. \times \ddot{\mathbf{\Phi}}_i^{[r]}(k)\mathbf{H}\mathbf{\Phi}^{[t]}(k)\mathbf{s}(k) \right\} \right]$$

$$= \frac{2}{\sigma_w^2} \mathbb{E}_{\boldsymbol{\phi}|\mathbf{s}} \left[\Re\left\{ \sum_{\bar{m}=1}^{N_t} \left([\mathbf{H}]_{1:N_r,\bar{m}} s_{\bar{m}}(k)e^{j\phi_{\bar{m}}^{[t]}(k)}\right)^H \mathbf{V}_i \right.\right.$$
$$\left.\left. \times \sum_{\bar{m}=1}^{N_t} \left([\mathbf{H}]_{1:N_r,\bar{m}} s_{\bar{m}}(k)e^{j\phi_{\bar{m}}^{[t]}(k)}\right) \right\} \right]$$

$$= \frac{2}{\sigma_w^2} \sum_{\bar{m}=1}^{N_t} |h_{i,\bar{m}}|^2 |s_{\bar{m}}(k)|^2, \ell = \bar{\ell} = N_t, \ldots, N,$$
$$i = \ell - (N_t - 1) \tag{A.12}$$

$$[\mathbf{\Pi}(k)]_{\ell,\bar{\ell}} = \frac{2}{\sigma_w^2} \mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\phi}|\mathbf{s}} \left[-\Re\left\{ \mathbf{y}^H(k)\dot{\mathbf{\Phi}}_i^{[r]}(k)\mathbf{H}\dot{\mathbf{\Phi}}_\ell^{[t]}(k)\mathbf{s}(k) \right\} \right]$$

$$= \frac{2}{\sigma_w^2} \mathbb{E}_{\boldsymbol{\phi}|\mathbf{s}} \left[\Re\left\{ \sum_{m=1}^{N_t} \left([\mathbf{H}]_{1:N_r,m} s_m(k)e^{j\phi_m^{[t]}(k)}\right)^H \right.\right.$$
$$\left.\left. \times \mathbf{V}_i [\mathbf{H}]_{1:N_r,\ell} s_\ell(k)e^{j\phi_\ell^{[t]}(k)} \right\} \right]$$

$$= \frac{2}{\sigma_w^2} |h_{i,l}|^2 |s_l(k)|^2,$$
$$\ell = 1, \ldots, N_t - 1, \bar{\ell} = N_t, \ldots, N, i = \bar{\ell} - (N_t - 1) \tag{A.13}$$

where in (A.12) and (A.13), $\mathbf{V}_i$ is an $N_r \times N_r$ matrix with all elements equal to 0 except its ith diagonal element, $[\mathbf{V}]_{i,i} = 1$. In addition, the last equality in (A.11) follows since $\mathbb{E}_{\boldsymbol{\phi}}\left[e^{j(\phi_\ell^{[t]}(k)-\phi_{\bar{\ell}}^{[t]}(k))}\right] = \mathbb{E}_{\boldsymbol{\phi}}\left[e^{j(\phi_{\bar{\ell}}^{[t]}(k)-\phi_\ell^{[t]}(k))}\right] = 0$. Using similar steps as in (A.13), we can obtain $[\mathbf{\Pi}(k)]_{\ell,\bar{\ell}} = \frac{2}{\sigma_w^2}|h_{i,\bar{l}}|^2|s_{\bar{l}}(k)|^2$, for $\bar{\ell} = 1, \ldots, N_t - 1, \ell = N_t, \ldots, N$, and $i = \ell - (N_t - 1)$, and $[\mathbf{\Pi}(k)]_{\ell,\bar{\ell}} = 0$, for $\ell = N_t, \ldots, N$, $\bar{\ell} = N_t, \ldots, N$, and $\ell \neq \bar{\ell}$. Finally, using the above results, the entries of $\mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\phi}|\mathbf{s}}\left[-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)} \log p(\mathbf{y}(k)|\boldsymbol{\phi}(k),\mathbf{s}(k))\right]$ can be obtained as given in (9) in Section III-A.

B. Derivation of $\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi})\right]$

According to the Wiener phase noise model in (4), $-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi})$ can be expanded as

$$-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}) = -\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(1))$$
$$- \sum_{k=2}^{K} \Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)). \tag{A.14}$$

Moreover, $p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))$ is given by

$$p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)) = \frac{1}{(2\pi)^{N/2}(\det \mathbf{\Sigma})^{1/2}}$$
$$\exp\left\{ -\frac{[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]^H \mathbf{\Sigma}^{-1}[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]}{2} \right\}. \tag{A.15}$$

Based on (A.14) and (A.15), the $KN \times KN$ matrix $\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi})\right]$ can be written as

$$\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi})\right] = \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(1))\right]$$
$$+ \sum_{k=2}^{K} \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))\right]. \tag{A.16}$$

Assuming no prior information about $\boldsymbol{\phi}(1)$, $\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(1))\right] = \mathbf{0}_{KN \times KN}$. Thus, (A.16) can be rewritten as

$$\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi})\right] = \sum_{k=2}^{K} \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))\right]. \tag{A.17}$$

Based on the definition of the Hessian operator, the $KN \times KN$ matrices, $\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))\right]$, for $k = 2, \ldots, K$, are zero everywhere except from the $((k-2)N+1)$th to the kNth row and the $((k-2)N+1)$th to the kNth column. The non-zero elements of $\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))\right]$, for $k = 2, \ldots, K$, are given by (A.18) at the bottom of this page. Note that (A.18) follows from the definition of $p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))$ in (A.15). Let us define $\gamma \triangleq \frac{1}{2}[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]^H \mathbf{\Sigma}^{-1}[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]$. The entries of the $N \times N$ matrix, $\mathbb{E}_{\boldsymbol{\phi}}\left[\Delta_{\boldsymbol{\phi}(k-1)}^{\boldsymbol{\phi}(k-1)}\gamma\right]$, for $m = 1, \ldots, N_t - 1$, $\tilde{m} = 1, \ldots, N_t - 1$, $n = 1, \ldots, N_r$, and $\tilde{n} = 1, \ldots, N_r$ can be determined as

$$\left[\mathbb{E}_{\boldsymbol{\phi}}\left[\Delta_{\boldsymbol{\phi}(k-1)}^{\boldsymbol{\phi}(k-1)}\gamma\right]\right]_{m,\tilde{m}}$$
$$= \mathbb{E}_{\boldsymbol{\phi}}\left[\frac{\partial^2 \gamma}{\partial\phi_m^{[t]}(k-1)\partial\phi_{\tilde{m}}^{[t]}(k-1)}\right]$$
$$= [\mathbf{\Sigma}^{-1}]_{m,\tilde{m}}, \tag{A.19a}$$

$$\left[\mathbb{E}_{\boldsymbol{\phi}}\left[\Delta_{\boldsymbol{\phi}(k-1)}^{\boldsymbol{\phi}(k-1)}\gamma\right]\right]_{\substack{n+N_t-1,\\ \tilde{n}+N_t-1}}$$
$$= \mathbb{E}_{\boldsymbol{\phi}}\left[\frac{\partial^2 \gamma}{\partial\phi_n^{[r]}(k-1)\partial\phi_{\tilde{n}}^{[r]}(k-1)}\right]$$
$$= [\mathbf{\Sigma}^{-1}]_{n+N_t-1,\tilde{n}+N_t-1}, \tag{A.19b}$$

$$\left[\mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))\right]\right]_{\substack{(k-2)N+1:(k-1)N,\\ (k-2)N+1:(k-1)N}} = \mathbb{E}_{\boldsymbol{\phi}}\left[-\Delta_{\boldsymbol{\phi}(k-1)}^{\boldsymbol{\phi}(k-1)} \log p(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1))\right]$$
$$= \mathbb{E}_{\boldsymbol{\phi}}\left[\Delta_{\boldsymbol{\phi}(k-1)}^{\boldsymbol{\phi}(k-1)}\left[\frac{1}{2}[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]^H \mathbf{\Sigma}^{-1}[\boldsymbol{\phi}(k)-\boldsymbol{\phi}(k-1)]\right]\right], \tag{A.18}$$

$$\left[\mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right] \right]_{\substack{(k-1)N+1:kN, \\ (k-1)N+1:kN}} = \mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k)} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right] = \boldsymbol{\Sigma}^{-1}, \tag{A.21}$$

$$\left[\mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right] \right]_{\substack{(k-2)N+1:(k-1)N, \\ (k-1)N+1:kN}} = \mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}(k-1)}^{\boldsymbol{\phi}(k)} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right] = -\boldsymbol{\Sigma}^{-1}, \tag{A.22}$$

$$\left[\mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right] \right]_{\substack{(k-1)N+1:kN, \\ (k-2)N+1:(k-1)N}} = \mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}(k)}^{\boldsymbol{\phi}(k-1)} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right] = -\boldsymbol{\Sigma}^{-1}. \tag{A.23}$$

$$\left[\mathbb{E}_{\boldsymbol{\phi}} \left[\Delta_{\boldsymbol{\phi}(k-1)}^{\boldsymbol{\phi}(k-1)} \gamma \right] \right]_{m,n+N_t-1}$$

$$= \mathbb{E}_{\boldsymbol{\phi}} \left[\frac{\partial \gamma}{\partial \phi_m^{[t]}(k-1) \partial \phi_n^{[r]}(k-1)} \right]$$

$$= \left[\boldsymbol{\Sigma}^{-1} \right]_{m,n+N_t-1}. \tag{A.19c}$$

In (A.19), $\boldsymbol{\Sigma}^{-1}$ can be calculated by applying the matrix inversion Lemma in [55], as shown in (10). Using (A.18) and (A.19), we also have

$$\left[\mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right] \right]_{\substack{(k-2)N+1:(k-1)N, \\ (k-2)N+1:(k-1)N}}$$
$$= \boldsymbol{\Sigma}^{-1}. \tag{A.20}$$

Applying similar steps as above, the remaining non-zero elements of $\mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p\big(\boldsymbol{\phi}(k)|\boldsymbol{\phi}(k-1)\big) \right]$, for $k = 2, \ldots, K$, are given by (A.21)–(A.23) at the top of the page. Consequently, using (A.20)–(A.23), $\mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}) \right]$ can be obtained as

$$\mathbb{E}_{\boldsymbol{\phi}} \left[-\Delta_{\boldsymbol{\phi}}^{\boldsymbol{\phi}} \log p(\boldsymbol{\phi}) \right]$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & -\boldsymbol{\Sigma}^{-1} & 0 & \ldots & 0 \\ -\boldsymbol{\Sigma}^{-1} & 2\boldsymbol{\Sigma}^{-1} & -\boldsymbol{\Sigma}^{-1} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\boldsymbol{\Sigma}^{-1} & 2\boldsymbol{\Sigma}^{-1} & -\boldsymbol{\Sigma}^{-1} \\ 0 & \ldots & 0 & -\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Sigma}^{-1}. \end{bmatrix}. \tag{A.24}$$

Finally, using (6), (9), and (A.24), the BIM can be derived in closed-form as shown in (7) in Section III-A.

REFERENCES

[1] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple antenna techniques for wireless communications—A comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 2, pp. 87–105, 2009.

[2] D. Gesbert, M. Shafi, D. S. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281–302, 2003.

[3] M. Dohler, R. W. Heath, A. Lozano, C. B. Papadias, and R. A. Valenzuela, "Is the PHY layer dead?," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 159–165, Apr. 2011.

[4] A. Chorti and M. Brookes, "A spectral model for RF oscillators with power-law phase noise," *IEEE Trans. Circuits Syst.*, vol. 53, no. 9, pp. 1989–1999, Sep. 2006.

[5] A. Hajimiri, S. Limotyrakis, and T. H. Lee, "Jitter and phase noise in ring oscillators," *IEEE J. Solid-State Circuits*, vol. 34, no. 6, pp. 790–804, Jun. 1999.

[6] L. Tomba, "On the effect of Wiener phase noise in OFDM systems," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 580–583, May 1998.

[7] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterization," *IEEE Trans. Circuits Syst.*, vol. 47, no. 5, pp. 655–674, May 2000.

[8] D. Baum and H. Bolcskei, "Impact of phase noise on MIMO channel measurement accuracy," in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Sep. 2004, vol. 3, pp. 1614–1618.

[9] P. Almers, S. Wyne, F. Tufvesson, and A. F. Molisch, "Effect of random walk phase noise on MIMO measurements," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, 2005, vol. 1, pp. 141–145.

[10] H. Mehrpouyan, A. A. Nasir, T. Eriksson, S. D. Blostein, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012.

[11] T. Höhne and V. Ranki, "Phase noise in beamforming," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 1696–1705, Dec. 2010.

[12] A. Taparugssanagorn and J. Ylitalo, "Characteristics of short-term phase noise of MIMO channel sounding and its effect on capacity estimation," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 1, pp. 196–201, Jan. 2009.

[13] L. Zhao and W. Namgoong, "A novel phase-noise compensation scheme for communication receivers," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 532–542, Mar. 2006.

[14] V. Simon, A. Senst, M. Speth, and H. Meyr, "Phase noise estimation via adapted interpolation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2001, vol. 6, pp. 3297–3301.

[15] N. Hadaschik, M. Dorpinghaus, A. Senst, O. Harmjanz, U. Kaufer, G. Ascheid, and H. Meyr, "Improving MIMO phase noise estimation by exploiting spatial correlations," in *Proc. IEEE Acous., Speech, Signal Process. (ICASSP)*, Mar. 2005, vol. 3, pp. 833–836.

[16] T. Ingason, H. Liu, M. Coldrey, A. Wolfgang, and J. Hansryd, "Impact of frequency selective channels on a line-of-sight MIMO microwave radio link," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2010, pp. 1–5.

[17] F. Bøhagen, P. Orten, and G. E. Øien, "Design of optimal high-rank line-of-sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 1420–1425, Apr. 2007.

[18] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York, NY, USA: Wiley-InterScience, 1997.

[19] S. A. Merritt, "The iterated extended Kalman phase detector," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 522–526, May 1989.

[20] J. Bhatti and M. Moeneclaey, "Feedforward data-aided phase noise estimation from a DCT basis expansion," *EURASIP J. Wireless Commun. Netw. (Special Issue on Synchronization in Wireless Communications)*, vol. 2009, Jan. 2009, Article ID 568570, doi: 10.1155/2009/568570, Article No. 4 [Online]. Available: http://dl.acm.org/citation.cfm?id=1661320

[21] J. Dauwels and H.-A. Loeliger, "Joint decoding and phase estimation: An exercise in factor graphs," in *Proc. IEEE Symp. Inf. Theory*, Jul. 2003, pp. 231–231.

[22] P. Amblard, J. Brossier, and E. Moisan, "Phase tracking: What do we gain from optimality? Particle filtering versus phase-locked loops," *Signal Process.*, vol. 83, no. 1, pp. 151–167, 2003.

[23] S. Bay, C. Herzet, J.-M. Brossier, J.-P. Barbot, and B. Geller, "Analytic and asymptotic analysis of Bayesian Cramér-Rao bound for dynamical phase offset estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 61–70, Jan. 2008.

[24] N. Noels, H. Steendam, M. Moeneclaey, and H. Bruneel, "Carrier phase and frequency estimation for pilot-symbol assisted transmission: Bounds and algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4578–4587, Dec. 2005.

[25] S. Godtmann, N. Hadaschik, A. Pollok, G. Ascheid, and H. Meyr, "Iterative code-aided phase noise synchronization based on the LMMSE criterion," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2007, pp. 1–5.

[26] T. S. Shehata and M. El-Tanany, "Joint iterative detection and phase noise estimation algorithms using Kalman filtering," in *Proc. Canad. Workshop Inf. Theory*, May 2009, pp. 165–168.

[27] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005.

[28] J. Bhatti and M. Moeneclaey, "Performance analysis of iterative decision-directed phase noise estimation," *Proc. Future Netw. Mobile Summit*, pp. 1–8, Jun. 2010.

[29] N. Noels, V. Lottici, A. Dejonghe, H. S. M. Moeneclaey, M. Luise, and L. Vandendorpe, "A theoretical framework for soft-information-based synchronization in iterative (turbo) receivers," *EURASIP J. Wireless Commun. Netw. (Special Issue on Advanced Signal Process. Algorithms for Wireless Communications)*, vol. 2005, no. 2, pp. 117–129, Apr. 2005.

[30] M. Nissilä and S. Pasupathy, "Adaptive iterative detectors for phase-uncertain channels via variational bounding," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 715–716, Mar. 2009.

[31] T. C. W. Schenk, X.-J. Tao, P. F. M. Smulders, and E. R. Fledderus, "Influence and suppression of phase noise in multi-antenna OFDM," in *Proc. 60th IEEE Veh. Technol. Conf. (VTC)—Fall*, Sep. 2004, vol. 2, pp. 1443–1447.

[32] T. C. W. Schenk, X.-J. Tao, P. F. M. Smulders, and E. R. Fledderus, "On the influence of phase noise induced ICI in MIMO OFDM systems," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 682–684, Aug. 2005.

[33] A. G. Armada, "Understanding the effects of phase noise in orthogonal frequency division multiplexing (OFDM)," *IEEE Trans. Broadcast.*, vol. 47, no. 2, pp. 153–159, Jun. 2001.

[34] A. G. Armada and M. Calvo, "Phase noise and sub-carrier spacing effects on the performance of an OFDM communication system," *IEEE Commun. Lett.*, vol. 2, no. 1, pp. 11–13, Jan. 1998.

[35] X. Zhang and H.-G. Ryu, "Joint estimation and suppression of phase noise and carrier frequency offset in multiple-input multiple-output single carrier frequency division multiple access with single-carrier space frequency block coding," *IET Trans. Commun.*, vol. 4, no. 16, pp. 1998–2007, 2010.

[36] Y. Wang and D. Falconer, "Phase noise estimation and suppression for single carrier SDMA uplink," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Jul. 2010, pp. 1–6.

[37] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Performance analysis of OFDM with Wiener phase noise and frequnecy selective fading channels," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1321–1331, May 2011.

[38] D. D. Lin, R. Pacheco, T. J. Lim, and D. Hatzinakos, "Joint estimation of channel response, frequency offset, and phase noise in OFDM," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3442–3554, Sep. 2006.

[39] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1862–1874, May 2007.

[40] T. Pollet, M. V. Blade, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 191–193, Feb./Mar./Apr. 1995.

[41] J. Wells, *Multi-Gigabit Microwave and Millimeter-Wave Wireless Communications*, 1st ed.   Norwood, MA, USA: Artech House, 2010.

[42] J. Hansryd and J. Edstam, "Microwave capacity evolution: New technologies and additional frequencies," in *Ericsson AB Publications*, 2011 [Online]. Available: http://www.ericsson.com/news/110621_microwave_capacity_evolution_244188810_c

[43] N. A. D'Andrea, U. Mengali, and R. Reggiannini, "Comparison of carrier recovery methods for narrow-band polyphase shift keyed signals," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 1988, vol. 3, pp. 1474–1478.

[44] M. Moeneclaey and G. de Jonghe, "ML-oriented NDA carrier synchronization for general rotationally symmetric signal constellations," *IEEE Trans. Commun.*, vol. 42, no. 8, pp. 2531–2533, Aug. 1994.

[45] B. G. Quinn, "Estimating frequency by interpolation using Fourier coefficients," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1264–1268, May 1994.

[46] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*.   Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[47] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.

[48] G. Klimovitch, "A nonlinear theory of near-carrier phase noise in free-running oscillators," in *Proc. IEEE Int. Conf. Circuits Syst.*, Mar. 2000, pp. 1–6.

[49] F. Herzel, "An analytical model for the power spectral density of a voltage controlled oscillator and its analogy to the laser linewidth theory," *IEEE Trans. Circuits Syst. I, Fund. Theory Appl.*, vol. 45, no. 9, pp. 904–908, Sep. 1998.

[50] G. Niu, "Noise in SIGe HBT RF technology: Physics, modeling, and circuit implications," *Proc. IEEE*, vol. 93, no. 9, pp. 1583–1597, Sep. 2005.

[51] J. Dauwels, "Computing Bayesian Cramer-Rao bounds," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Australia, Sep. 2005, pp. 425–429.

[52] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*.   New York, NY, USA: Wiley, 2001.

[53] J. Jain, H. Li, S. Cauley, C.-K. Koh, and V. Balakrishnan, "Numerically stable algorithms for inversion of block tridiagonal and banded matrices," Purdue Univ., Lafayette, IN, USA, ECE Tech. Rep., Jun. 2007 [Online]. Available: http://docs.lib.purdue.edu/ecetr/357

[54] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, 1998.

[55] R. A. Horn and C. R. Johnson, *Matrix Analysis*.   Cambridge, U.K.: Cambridge Univ. Press, 1990.

[56] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 602–613, Mar. 2003.

[57] S. Song, A. C. Singer, and K.-M. Sung, "Soft input channel estimation for turbo equalization," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2885–2894, Oct. 2004.

[58] I. Ziskand and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[59] J. B. Fraleigh and R. A. Beauregard, *Linear Algebra*.   New York, NY, USA: Addison-Wesley, 1987.

[60] G. L. Stüber, *Principles of Mobile Communication*, Second ed.   Norwell, MA, USA: Kluwer, 2001.

[61] H. Mehrpouyan and S. D. Blostein, "Bounds and algorithms for multiple frequency offset estimation in cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1300–1311, Apr. 2011.

[62] P. Z. Peebles, *Probability, Random Variables and Random Signal Principles*.   New York, NY, USA: McGraw-Hill, 2001.

Ali A. Nasir (S'11) received the B.Sc. (First Class Hons.) degree in electrical engineering from the University of Engineering and Technology (UET), Lahore, Pakistan, in 2007. He is currently working towards the Ph.D. degree at the Research School of Engineering, the Australian National University (ANU), Canberra, Australia.

Prior to joining ANU, he worked as a Lecturer in UET for nine months and then as a Design Engineer in the Centre for Advanced Research in Engineering (CARE), Islamabad, Pakistan, for a year. He was a research visitor in Chalmers University of Technology, Gothenburg, Sweden, for six months from April until October 2011. His research interests are in the area of synchronization and channel estimation in cooperative communication, MIMO, and OFDM systems.

Mr. Nasir was awarded a University Gold Medal for outstanding performance during the final year of his undergraduate studies. He is a recipient of an ANU International Ph.D. scholarship for the duration of his Ph.D. He was also awarded an ANU Vice-Chancellor's Higher Degree Research (HDR) travel grant in 2011.

Hani Mehrpouyan (S'05–M'10) received the B.Sc. (Hons.) degree in computer engineering from Simon Fraser University, Burnaby, Canada, in 2004. In 2005, he joined the Department of Electrical and Computer Engineering at Queen's University, Kingston, Canada, as a Ph.D. candidate, where he received the Ph.D. degree in Electrical Engineering in 2010.

From September 2010 to March of 2012, he was a Postdoctoral Researcher at the Department of Signal and Systems at Chalmers University of Technology,

where he led the MIMO aspects of the microwave backhauling for next-generation wireless networks project. From April 2012 to August 2012, he served as a Research Associate at the University of Luxembourg, where was responsible for new interference cancellation and synchronization schemes for next-generation satellite communication links. Since August 2012, he has been an Assistant Professor at the Department of Computer and Electrical Engineering and Computer Science at California State University, Bakersfield, CA, USA. He has more than 25 publications in prestigious IEEE journals and conferences. He has also been involved with industry leaders, such as Ericsson AB, Research in Motion (RIM), and Alcatel. His current research interests lie in the area of applied signal processing and physical layer of wireless communication systems, including synchronization, channel estimation, interference cancelation, and performance optimization.

Dr. Mehrpouyan has received more than ten scholarships and awards, e.g., the IEEE WIRELESS COMMUNICATION LETTERS Exemplary Reviewer Award, the IEEE Globecom Early Bird Student Award, the NSERC-IRDF, NSERC PGS-D, NSERC CGS-M Alexander Graham Bell Award, and the B.C. Wireless Innovation Award. He has served as a TPC member for the IEEE Globecom, VTC, among others. For more information, refer to www.mehrpouyan.info.

2004 Innovations Award of the Vodafone Foundation for Research in Mobile Communications, the 2006 UBC Killam Research Prize, the 2007 Wilhelm Friedrich Bessel Research Award of the Alexander von Humboldt Foundation, the 2008 Charles McDowell Award for Excellence in Research from UBC, a 2011 Alexander von Humboldt Professorship, and a 2012 NSERC E.W.R. Steacie Fellowship. In addition, he received Best Paper Awards from the German Information Technology Society (ITG); the European Association for Signal, Speech and Image Processing (EURASIP); the IEEE WCNC 2012; the IEEE Globecom 2011; the IEEE ICUWB 2006; the International Zurich Seminar on Broadband Communications; and the European Wireless 2000. He is a Fellow of the Canadian Academy of Engineering and a Fellow of the Engineering Institute of Canada. He is currently the Editor-in-Chief of the IEEE TRANSACTIONS ON COMMUNICATIONS.

Robert Schober (S'98–M'01–SM'08–F'10) was born in Neuendettelsau, Germany, in 1971. He received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from the University of Erlangen-Nuermberg, Germany, in 1997 and 2000, respectively.

From May 2001 to April 2002, he was a Postdoctoral Fellow at the University of Toronto, ON, Canada, sponsored by the German Academic Exchange Service (DAAD). Since May 2002, he has been with the University of British Columbia (UBC), Vancouver, Canada, where he is currently a Full Professor. Since January 2012, he has been an Alexander von Humboldt Professor and the Chair for Digital Communication at the Friedrich Alexander University (FAU), Erlangen, Germany. His research interests fall into the broad areas of communication theory, wireless communications, and statistical signal processing.

Dr. Schober received several awards for his work, including the 2002 Heinz Maier-Leibnitz Award of the German Science Foundation (DFG), the

Yingbo Hua (S'86–M'88–SM'92–F'02) received the B.S. degree from Southeast University, Nanjing, China, in 1982 and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, USA, in 1983 and 1988, respectively.

He held faculty positions with the University of Melbourne, Australia, from 1990 to 2001. He was on leave with the Hong Kong University of Science and Technology from 1999 to 2000, and consulting with Microsoft Research, WA, USA, in summer 2000. He has been a Professor with the University of California at Riverside, CA, USA, since 2001. His interests span the fields of sensing, signal processing and communications, where he has published hundreds of papers with more than 6000 citations.

Dr. Hua was elected to AAAS Fellow in 2011.