(article starts on next page)

Algorithms for Joint Phase Estimation and Decoding for MIMO Systems in the Presence of Phase Noise and Quasi-Static Fading Channels

Rajet Krishnan, *Student Member, IEEE,* Giulio Colavolpe, *Senior Member, IEEE*
Alexandre Graell i Amat, *Senior Member, IEEE,* and Thomas Eriksson

Abstract—In this work, we derive the maximum a posteriori (MAP) symbol detector for a multiple-input multiple-output system in the presence of Wiener phase noise due to noisy local oscillators, and quasi-static fading channels. As in single-antenna systems, the computation of the optimum receiver is an analytically intractable problem and is unimplementable in practice. In this purview, we propose three suboptimal, low-complexity algorithms for approximately implementing the MAP symbol detector, which involve joint phase noise estimation and data detection. Our first algorithm is obtained by means of the sum-product algorithm, where we use the multivariate Tikhonov canonical distribution approach. In our next algorithm, we derive an approximate MAP symbol detector based on the smoother-detector framework, wherein the detector is properly designed by incorporating the phase noise statistics from the smoother. The third algorithm is derived based on the variational Bayesian framework. By simulations, we evaluate the performance of the proposed algorithms for both uncoded and coded data transmissions, and we observe that the proposed techniques significantly outperform the other important algorithms from prior works, which are considered in this work.

Index Terms – Maximum a posteriori (MAP) detection, phase noise, sum-product algorithm (SPA), variational Bayesian (VB) framework, extended Kalman smoother (EKS), MIMO.

I. INTRODUCTION

EMPLOYING multiple-input multiple-output (MIMO) systems has been shown to significantly enhance performance in terms of data rate and link reliability in wireless fading environments [1]. In general, the analysis and design of MIMO system is based on the assumption that the carrier phase is perfectly known at the receiver, and that there is no phase noise in the system. The phase noise manifests in a MIMO system as the random, time-varying phase differences between the oscillators connected to the antennas at the transmitter and the receiver. Practical designs of MIMO systems based on this assumption can result in significant performance losses and have to be addressed appropriately [2]. The detrimental effects of phase noise can be even more pronounced in scenarios where independent oscillators are connected to each transmit and receive antenna (or a subset of them). This scenario is particularly relevant for line-of-sight MIMO systems that operate at carrier frequencies of around 10 GHz or lesser. Here, separate oscillators are needed

for each antenna [3], since the antennas are placed far from each other [4]. The scenario under consideration also corresponds to a massive MIMO system [5], [6], where a large number of antennas are placed at the base station and each user terminal is equipped with a single antenna.

The problem of designing receiver algorithms in the presence of random, time-varying phase noise due to noisy local oscillators has been studied extensively for single-antenna systems. We refer the readers to [7]–[10] and the references therein. To address the problem of designing receiver algorithms for joint phase noise estimation and data detection, the expectation-maximization (EM) framework is applied in [7], resulting in a code-aided synchronization technique. In [8], receiver algorithms are developed based on the sum-product algorithm (SPA) by constraining the probability density functions (pdfs) computed by the SPA to be in a certain canonical family (for e.g., the exponential family). This method of constraining the pdfs is referred to as the canonical distribution approach [11], and in particular, using the Tikhonov canonical distribution in [8] results to be the most convenient and effective choice. The variational Bayesian (VB) framework is adopted in [9] to develop efficient algorithms for joint phase noise estimation and data detection. In [10], receiver algorithms are derived by using a smoother-detector structure based on the maximum a posteriori (MAP) symbol detector derived in [12], where the detector is properly designed by incorporating the phase noise statistics from the smoother.

The effect of phase noise on MIMO systems has been investigated in some recent work [2], [6], [13], [14], where the impact of phase noise on the MIMO channel measurements and the estimated capacity is studied. In [3], data-aided estimation of phase noise is studied using a Wiener filter. In [13], the problem of joint channel and phase noise estimation in a MIMO system is explored, and bounds on the estimation performance are derived. Soft-symbol aided estimation using an extended Kalman Smoother (EKS) and relevant estimation bounds are investigated in [15]. However, these works do not consider the problem of designing receiver algorithms for joint phase noise estimation and data detection. One of the few works investigating this problem can be found in [16], where the VB framework is employed. In general, MIMO receiver design has focused on developing algorithms for joint channel estimation and data detection (refer to [19], [20] and the references therein)—it is perceived that the phase noise can be handled by existing channel estimation-data detection algorithms since it can be regarded to be a part of the channel [2].

In this paper, we consider the problem of designing receiver

Rajet Krishnan, Alexandre Graell i Amat, and Thomas Eriksson are with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden (e-mail: {rajet, alexandre.graell, thomase}@chalmers.se).

Giulio Colavolpe is with Dipartimento di Ingegneria dell'Informazione, University of Parma, Parma, Italy (e-mail: giulio@unipr.it).

algorithms for joint phase noise estimation and data detection in a MIMO system, where each transmit and receive antenna is connected to an independent noisy oscillator. We focus on the scenario where the phase noise process is a discrete Wiener process [18] and drifts much faster than the channel [6], [16]. This implies that the phase noise in the system cannot be handled by moving it into the channel matrix and then compensating it by means of channel estimation—this is a typical scenario when noisy oscillators are used in the system and the channel is essentially quasi-static fading. For instance, in the case of fixed access networks like microwave backhaul networks, where the channel essentially remains a constant for a long period of time (quasi-static fading), the phase noise changes much faster than the channel [17]. Separate phase noise estimators are also crucial in line-of-sight (LoS) MIMO systems, where a full-rank channel matrix is achieved by a careful placement of the antennas. In this case, the channel is almost constant and it varies much slower than phase noise. Finally, for the scenarios considered above, as the carrier frequency increases, the phase noise innovation variance will increase significantly, and the phase noise will be required to be tracked symbol-by-symbol.

For the MIMO system under consideration, we derive the MAP symbol detector which minimizes the symbol error probability. This receiver structure explicitly involves the estimation of the a posteriori pdf of the phase noise and data detection. The computation of the a posteriori phase noise pdf is analytically intractable for the Wiener phase noise process. This motivates the need for investigating practical, low complexity receiver algorithms for joint phase noise estimation and data detection that also have a good performance. To this end, we propose three new algorithms based on the sum-product algorithm (SPA), the smoother-detector framework from [10], [12], and the VB framework in [9], respectively, for arbitrary number of transmit and receive antennas. We evaluate the performance of the proposed algorithms in strong phase noise scenarios in the presence of Rayleigh fading. We consider both uncoded and coded data transmissions, and compare the performance of the proposed algorithms with that of those available in the literature. We observe that the proposed algorithms significantly outperform those from literature, which are considered in this work.

The remainder of the paper is as follows. In Section II, the MIMO system model under study is presented. We derive the optimal MAP symbol detector in Section III. In Sections IV, V, and VI, we derive the SPA-based, smoother-detector-based, and VB-based algorithms, respectively. The computational complexity of the receivers is studied in Section VII. We present our simulation results in Section VIII. Finally, we summarize our key findings in Section IX.

Notation: the expectation and variance operators are denoted as $\mathbb{E}[\cdot]$ and $\mathrm{Var}(\cdot)$, respectively. The conjugate of a complex number is denoted as $[\cdot]^*$, and $\jmath = \sqrt{-1}$. $\Re\{\cdot\}$, $\Im\{\cdot\}$, $|\cdot|$, and $\angle\cdot$ are the real, imaginary part, magnitude, and angle of a complex number, respectively. The pdf and probability mass function (pmf) of a random variable are denoted as $p(\cdot)$ and $P(\cdot)$, respectively.

II. SYSTEM MODEL

We consider a MIMO system with N_t transmit antennas and N_r receive antennas. Data is transmitted as frames consisting of L symbols, and we consider both coded and uncoded transmission. The channel between the transmit-receive antennas is assumed to be a constant over the length of a frame, and is known (or estimated). Each antenna is equipped with an independent free-running oscillator that is perturbed by a random phase noise process, which varies much faster than the channel [16], and conforms to the model in (1) and (2). Specifically, the phase noise process from the oscillators is assumed to be varying from symbol-to-symbol as in (2). We assume that for given frame, the channel and phase noise are first jointly estimated as in [13]. This is done by first transmitting a training sequence that is used to estimate the channel and the phase noise by means of a least square estimator. The training sequence is then followed by the transmission of data symbols, during which an autonomous phase noise estimation algorithm is used to track the time-varying phase noise followed by data detection. The joint channel and phase noise estimates obtained from the training sequence are used as the true channel values in the (autonomous) phase noise estimation and data detection algorithm.

Assuming square-root Nyquist pulses for transmission, and matched filtering followed by sampling at symbol period T_s, the received signal in the kth time instant at the nth receive antenna is

$$
\begin{aligned}
r_k^{(n)} &= \sum_{m=1}^{N_\mathrm{t}} h^{(m,n)} c_k^{(m)} e^{\jmath(\theta_{\mathrm{t},k}^{(m)} + \theta_{\mathrm{r},k}^{(n)})} + w_k^{(n)} \\
&\triangleq \sum_{m=1}^{N_\mathrm{t}} c_{\mathrm{h},k}^{(m,n)} e^{\jmath\theta_k^{(m,n)}} + w_k^{(n)},
\end{aligned}
\tag{1}
$$

where perfect timing and frequency synchronization is assumed [8]. In (1), $c_k^{(m)} \in \mathcal{M}$ is the symbol transmitted from the mth transmit antenna at the kth time instant and drawn equiprobably from an M-ary signal constellation set $\mathcal{M}$, $h^{(m,n)}$ represents the known (or estimated) channel realization between the mth transmit and nth receive antenna, $c_{\mathrm{h},k}^{(m,n)} \triangleq c_k^{(m)} h^{(m,n)}$, and $w_k^{(n)} \sim \mathcal{CN}(0, N_0)$ denotes the zero-mean additive white Gaussian noise (AWGN) at the nth receive antenna. The phase noise in the (m,n)th link, $\theta_k^{(m,n)}$, is defined as the sum of the discrete Wiener phase noise process from the oscillators connected to the mth transmit and the nth receive antenna, respectively, at time instant k, i.e., $\theta_k^{(m,n)} \triangleq \theta_{\mathrm{t},k}^{(m)} + \theta_{\mathrm{r},k}^{(n)}$, where

$$
\begin{aligned}
\theta_{\mathrm{t},k}^{(m)} &= \theta_{\mathrm{t},k-1}^{(m)} + \Delta_{\mathrm{t},k}^{(m)} \\
\theta_{\mathrm{r},k}^{(n)} &= \theta_{\mathrm{r},k-1}^{(n)} + \Delta_{\mathrm{r},k}^{(n)}.
\end{aligned}
\tag{2}
$$

In (2), $\Delta_{\mathrm{t},k}^{(m)} \sim \mathcal{N}(0, \sigma_\mathrm{t}^2)$, $\Delta_{\mathrm{r},k}^{(n)} \sim \mathcal{N}(0, \sigma_\mathrm{r}^2)$, and $\theta_{\mathrm{t},0}^{(m)}$ and $\theta_{\mathrm{r},0}^{(n)}$ are uniformly distributed in $[0, 2\pi)$. We remark that the values of σ_t^2 and σ_r^2 determine the quality of the oscillators used at the transmitter and the receiver, respectively.

Based on the received signal model in (1) and (2), we define the following vectors: $\mathbf{\Theta}_k \triangleq [\theta_{\mathrm{t},k}^{(1)}, \ldots, \theta_{\mathrm{t},k}^{(N_\mathrm{t})}, \theta_{\mathrm{r},k}^{(1)}, \ldots, \theta_{\mathrm{r},k}^{(N_\mathrm{r})}]^T$, $\bar{\mathbf{\Theta}} \triangleq [\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_L]^T$, $\mathbf{c}_k \triangleq [c_k^{(1)}, \ldots, c_k^{(N_\mathrm{t})}]^T$, $\bar{\mathbf{c}} \triangleq [\mathbf{c}_1, \ldots, \mathbf{c}_L]^T$, $\mathbf{r}_k \triangleq [r_k^{(1)}, \ldots, r_k^{(N_\mathrm{r})}]^T$, $\bar{\mathbf{r}} \triangleq [\mathbf{r}_1, \ldots, \mathbf{r}_L]^T$, $\mathbf{h} \triangleq [h^{(1,1)}, \ldots, h^{(N_\mathrm{t}, N_\mathrm{r})}]^T$, and $\mathbf{w}_k \triangleq [w_k^{(1)}, \ldots, w_k^{(N_\mathrm{r})}]^T$.

III. MAP SYMBOL DETECTOR

In this section, we derive the MAP symbol detector. Based on the received signal model in (1), the optimum receiver is

obtained as

$$\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k} \sum_{\bar{\mathbf{c}}\backslash\{\mathbf{c}_k\}} P(\bar{\mathbf{c}}|\bar{\mathbf{r}},\mathbf{h}) \tag{3}$$

$$= \arg\max_{\mathbf{c}_k} P(\mathbf{c}_k|\bar{\mathbf{r}},\mathbf{h}) \tag{4}$$

$$\propto \arg\max_{\mathbf{c}_k} P(\mathbf{c}_k)p(\bar{\mathbf{r}}|\mathbf{c}_k,\mathbf{h}) \tag{5}$$

$$= \arg\max_{\mathbf{c}_k} \int_{\boldsymbol{\Theta}_k} P(\mathbf{c}_k)p(\bar{\mathbf{r}}|\mathbf{c}_k,\boldsymbol{\Theta}_k,\mathbf{h})p(\boldsymbol{\Theta}_k|\mathbf{c}_k,\mathbf{h})\mathrm{d}\boldsymbol{\Theta}_k$$

$$= \arg\max_{\mathbf{c}_k} \int_{\boldsymbol{\Theta}_k} P(\mathbf{c}_k)p(\mathbf{r}_k|\mathbf{c}_k,\boldsymbol{\Theta}_k,\bar{\mathbf{r}}_k,\mathbf{h})p(\bar{\mathbf{r}}_k|\mathbf{c}_k,\boldsymbol{\Theta}_k,\mathbf{h})$$
$$\cdot\, p(\boldsymbol{\Theta}_k|\mathbf{c}_k,\mathbf{h})\mathrm{d}\boldsymbol{\Theta}_k \tag{6}$$

$$= \arg\max_{\mathbf{c}_k} \int_{\boldsymbol{\Theta}_k} P(\mathbf{c}_k)p(\mathbf{r}_k|\mathbf{c}_k,\boldsymbol{\Theta}_k,\mathbf{h})p(\bar{\mathbf{r}}_k|\mathbf{c}_k,\mathbf{h})$$
$$\cdot\, p(\boldsymbol{\Theta}_k|\mathbf{c}_k,\bar{\mathbf{r}}_k,\mathbf{h})\mathrm{d}\boldsymbol{\Theta}_k \tag{7}$$

$$\propto \arg\max_{\mathbf{c}_k} \int_{\boldsymbol{\Theta}_k} P(\mathbf{c}_k)p(\mathbf{r}_k|\mathbf{c}_k,\boldsymbol{\Theta}_k,\mathbf{h})p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}}_k)\mathrm{d}\boldsymbol{\Theta}_k. \tag{8}$$

In (3), we express the MAP symbol detector for the symbols transmitted in the kth time instant as the marginalization of the a posteriori pmf of $\bar{\mathbf{c}}$ with respect to all the symbols but $\mathbf{c}_k$. In (5), $P(\mathbf{c}_k)$ represents the a priori probability of the transmitted symbols in the kth time instant, and we apply that $\mathbf{h}$ is independent of $\mathbf{c}_k$. We define $\bar{\mathbf{r}}_k \triangleq [\mathbf{r}_1,\ldots,\mathbf{r}_{k-1},\mathbf{r}_{k+1},\ldots,\mathbf{r}_L]$ in (6). In (7), it is applied that, given $\mathbf{c}_k$, $\mathbf{h}$ and $\boldsymbol{\Theta}_k$, $\bar{\mathbf{r}}_k$ is independent of $\mathbf{r}_k$. It is assumed in (8) that $\mathbf{c}_k$ and $\bar{\mathbf{r}}_k$ are independent of each other, which holds in the case of uncoded data transmission, implying that we can omit $p(\bar{\mathbf{r}}_k|\mathbf{c}_k,\mathbf{h}) = p(\bar{\mathbf{r}}_k|\mathbf{h})$. Furthermore, we apply that $\mathbf{c}_k,\mathbf{h}$ and $\boldsymbol{\Theta}_k$ are independent of each other.

The detector obtained in (8) is a vector extension of the MAP symbol detector derived by Kam *et al.* in [12] – it detects $\mathbf{c}_k$ based on the conditional pdf of $\boldsymbol{\Theta}_k$, $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}}_k)$, which is computed using all received signals outside the kth time instant. The integral in (8) represents the a posteriori pmf of the transmitted symbols that is obtained after the marginalization of the phase noise. In uncoded systems, the transmitted symbols are detected based on (8). For the Wiener phase noise process, determining $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}}_k)$ is analytically intractable, which also makes the MAP detector intractable and unimplementable in practice [12].

The MAP detector presented in (8) can also be obtained by applying the SPA based on the factor graph framework [21]. This analysis forms the basis of the algorithm that is presented in Section IV. In order to derive the MAP detector using the SPA, we rewrite (3) as

$$\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k} \sum_{\bar{\mathbf{c}}\backslash\{\mathbf{c}_k\}} P(\bar{\mathbf{c}}|\bar{\mathbf{r}},\mathbf{h})$$
$$= \arg\max_{\mathbf{c}_k} \sum_{\bar{\mathbf{c}}\backslash\{\mathbf{c}_k\}} \int_{\bar{\boldsymbol{\Theta}}} P(\bar{\mathbf{c}},\bar{\boldsymbol{\Theta}}|\bar{\mathbf{r}},\mathbf{h})\mathrm{d}\bar{\boldsymbol{\Theta}}, \tag{9}$$

Factorizing the integrand, we obtain

$$P(\bar{\mathbf{c}},\bar{\boldsymbol{\Theta}}|\bar{\mathbf{r}},\mathbf{h}) \propto P(\bar{\mathbf{c}})p(\bar{\boldsymbol{\Theta}}|\bar{\mathbf{c}})p(\bar{\mathbf{r}}|\bar{\mathbf{c}},\bar{\boldsymbol{\Theta}},\mathbf{h}),$$
$$= P(\boldsymbol{\Theta}_0)\prod_{k=1}^{L} P(\mathbf{c}_k)\underbrace{p(\boldsymbol{\Theta}_k|\boldsymbol{\Theta}_{k-1})}_{p_\Delta(\boldsymbol{\Theta}_k-\boldsymbol{\Theta}_{k-1})}p(\mathbf{r}_k|\boldsymbol{\Theta}_k,\mathbf{c}_k,\mathbf{h}). \tag{10}$$

To factorize the function in (10) we exploit the fact that $\boldsymbol{\Theta}_k$ is a discrete Wiener process as in (2). Furthermore, in the sequel, we refrain from conditioning with respect to $\mathbf{h}$ since it is known,.

The FG associated with (10) is drawn in Fig. 1. With reference to the messages in the figure, we have

$$P_{\mathrm{d}}^{(c)}(\mathbf{c}_k) = P(\mathbf{c}_k) \tag{11}$$

$$p_{\mathrm{d}}^{(\theta)}(\boldsymbol{\Theta}_k) = \sum_{\mathbf{c}_k} P_{\mathrm{d}}^{(c)}(\mathbf{c}_k)p(\mathbf{r}_k|\mathbf{c}_k,\boldsymbol{\Theta}_k) \tag{12}$$

$$p_{\mathrm{f}}^{(\theta)}(\boldsymbol{\Theta}_k) = \int_{\boldsymbol{\Theta}_{k-1}} p_{\mathrm{f}}^{(\theta)}(\boldsymbol{\Theta}_{k-1})p_{\mathrm{d}}^{(\theta)}(\boldsymbol{\Theta}_{k-1})$$
$$\cdot\, p_\Delta(\boldsymbol{\Theta}_k-\boldsymbol{\Theta}_{k-1})\mathrm{d}\boldsymbol{\Theta}_{k-1} \tag{13}$$

$$p_{\mathrm{b}}^{(\theta)}(\boldsymbol{\Theta}_k) = \int_{\boldsymbol{\Theta}_{k+1}} p_{\mathrm{b}}^{(\theta)}(\boldsymbol{\Theta}_{k+1})p_{\mathrm{d}}^{(\theta)}(\boldsymbol{\Theta}_{k+1})$$
$$\cdot\, p_\Delta(\boldsymbol{\Theta}_{k+1}-\boldsymbol{\Theta}_k)\mathrm{d}\boldsymbol{\Theta}_{k+1} \tag{14}$$

$$P_{\mathrm{u}}^{(c)}(\mathbf{c}_k) = \int_{\boldsymbol{\Theta}_k} p_{\mathrm{f}}^{(\theta)}(\boldsymbol{\Theta}_k)p_{\mathrm{b}}^{(\theta)}(\boldsymbol{\Theta}_k)p(\mathbf{r}_k|\mathbf{c}_k,\boldsymbol{\Theta}_k)\mathrm{d}\boldsymbol{\Theta}_k. \tag{15}$$

Note that in the case of uncoded transmission the FG in Fig. 1 is a tree, and thus applying the SPA on this graph renders the exact MAP symbol detector (8). In this view, $p_{\mathrm{b}}^{(\theta)}(\boldsymbol{\Theta}_k)p_{\mathrm{f}}^{(\theta)}(\boldsymbol{\Theta}_k)$ is equal to the a posteriori pdf $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}}_k)$ in (8). Thus, the detector in (8) can be expressed in terms of $P_{\mathrm{u}}^{(c)}(\mathbf{c}_k)$ as

$$\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k} P_{\mathrm{u}}^{(c)}(\mathbf{c}_k)P_{\mathrm{d}}^{(c)}(\mathbf{c}_k)$$
$$\propto \arg\max_{\mathbf{c}_k} P_{\mathrm{u}}^{(c)}(\mathbf{c}_k), \tag{16}$$

where (16) simplifies since $P(\mathbf{c}_k)$ is uniform in an uncoded transmission. In coded systems, the same messages in (11)-(15) will be used, but now the message $P_{\mathrm{d}}^{(c)}(\mathbf{c}_k)$ is not the a priori pmf $P(\mathbf{c}_k)$, but rather the extrinsic symbol pmf provided by the decoder. Also, the message $P_{\mathrm{u}}^{(c)}(\mathbf{c}_k)$ in (15) is used for computing the bit log-likelihood ratios (LLRs) for soft decoding [10].

The messages in (11)-(15) form the core of the SPA for the implementation of the MAP detector. The implementation of the exact SPA is impractical because it involves the computation of the continuous pdfs of $\boldsymbol{\Theta}_k$ in (11)-(14), which are analytically intractable. The intractability of the exact MAP symbol detector in (8) and (16) motivates the need to explore practical, low complexity receiver algorithms. In the sequel, we present three receiver algorithms that are approximations of the exact MAP detector presented in (4), (8) and (16).

IV. MULTIVARIATE TIKHONOV-PARAMETERIZATION BASED SUM-PRODUCT ALGORITHM FOR APPROXIMATE MAP DETECTION

In the following, we derive a low-complexity SPA for the approximate implementation of the MAP symbol detector based on the canonical distribution approach suggested in [11]. This approach involves constraining the messages on the FG to a specific family of pdfs that can compactly and completely be described by a finite number of parameters. Thus, the task of computing the exact pdf is reduced to computing the parameters of the pdf. More specifically, we adopt the Tikhonov canonical distribution approach introduced by Colavolpe *et al.* in [8]; we constrain $p_{\mathrm{f}}^{(\theta)}(\boldsymbol{\Theta}_k)$ and $p_{\mathrm{b}}^{(\theta)}(\boldsymbol{\Theta}_k)$ to be multivariate Tikhonov pdfs in order to obtain a practical algorithm with good performance.
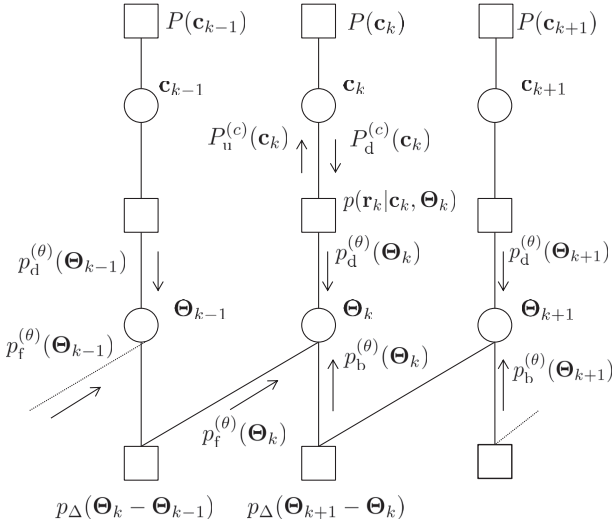
Fig. 1: Factor Graph and the SPA messages based on (10)

Without loss of generality we consider the case where $N_t = 2$ and $N_r = 1$, hence $\boldsymbol{\Theta}_k = [\theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)}]$ and $\mathbf{c}_k = [c_k^{(1)}, c_k^{(2)}]$. The generalization of the algorithm to arbitrary values of N_t and N_r is straightforward and is presented in Section IV-D. The received signal model in the kth time instant is

$$r_k^{(1)} = c_{h,k}^{(1,1)} e^{j(\theta_{t,k}^{(1)}+\theta_{r,k}^{(1)})} + c_{h,k}^{(2,1)} e^{j(\theta_{t,k}^{(2)}+\theta_{r,k}^{(1)})} + w_k^{(1)}. \quad (17)$$

Then (10) can be expressed as

$$P(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}|\bar{\mathbf{r}}) \propto P(\bar{\mathbf{c}})p(\theta_{t,0}^{(1)}, \theta_{t,0}^{(2)}, \theta_{r,0}^{(1)}) \prod_k p_\Delta(\boldsymbol{\Theta}_k - \boldsymbol{\Theta}_{k-1})$$
$$\prod_k p(r_k^{(1)}|c_k^{(1)}, c_k^{(2)}, \theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)}), \quad (18)$$

where

$$p(r_k^{(1)}|c_k^{(1)}, c_k^{(2)}, \theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)})$$
$$\propto \exp\left\{ -\frac{\left| r_k^{(1)} - c_{h,k}^{(1)} e^{j(\theta_{t,k}^{(1)}+\theta_{r,k}^{(1)})} - c_{h,k}^{(2)} e^{j(\theta_{t,k}^{(2)}+\theta_{r,k}^{(1)})} \right|^2}{N_0} \right\}. \quad (19)$$

We first seek to determine the functional form of the message $p_d^{(\theta)}(\boldsymbol{\Theta}_k)$ which is used to determine the messages $p_f^{(\theta)}(\boldsymbol{\Theta}_k)$ and $p_b^{(\theta)}(\boldsymbol{\Theta}_k)$. From (12),

$$p_d^{(\theta)}(\boldsymbol{\Theta}_k) = \sum_{c_k^{(1)}} \sum_{c_k^{(2)}} P_d^{(c)}(c_k^{(1)}, c_k^{(2)}) p(r_k^{(1)}|c_k^{(1)}, c_k^{(2)}, \theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)})$$
$$= p(r_k|\theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)}). \quad (20)$$

As we shall see in the sequel, $p_d^{(\theta)}(\boldsymbol{\Theta}_k)$ is a mixture of Tikhonov pdfs in $\theta_{t,k}^{(m)}, \theta_{r,k}^{(n)}$, or a mixture of Gaussian pdfs in $r_k^{(1)}$ when the transmitted symbol is unknown to the receiver. Deriving a low complexity algorithm mandates the approximation of this mixture. As in [8], [19], we approximate $p_d^{(\theta)}(\boldsymbol{\Theta}_k)$ as a single mode pdf in $r_k^{(1)}$. Specifically, we approximate $p(r_k|\theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)})$ by the Gaussian pdf that is closest in terms of the Kullback Leibler (KL) divergence measure. This is achieved by moment matching, since the Gaussian pdf belongs to the exponential family of distributions [22]. The meaningfulness

of this approximation stems from the fact that in many interesting scenarios, the sum of Gaussians constituting $p_d^{(\theta)}(\boldsymbol{\Theta}_k)$ has a dominant term. This is particularly true when the a priori information about the transmitted symbols is reliable as in coded systems [23]. In this case, $p_d^{(\theta)}(\boldsymbol{\Theta}_k)$ can have a pronounced mode or peak. However, as the size of the constellation is increased or the pilot density is reduced, this approximation can be lossy [24].

The mean and variance of the closest Gaussian pdf are

$$\mathbb{E}\{r_k^{(1)}|\theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)}\} = \alpha_k^{(1,1)} e^{j(\theta_{t,k}^{(1)}+\theta_{r,k}^{(1)})} + \alpha_k^{(2,1)} e^{j(\theta_{t,k}^{(2)}+\theta_{r,k}^{(1)})}$$

$$\text{Var}\{r_k^{(1)}|\theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)}\} = \beta_k^{(1,1)} + \beta_k^{(2,1)} + N_0 - \left|\alpha_k^{(1,1)}\right|^2$$
$$- \left|\alpha_k^{(2,1)}\right|^2 \triangleq \gamma_k^{(1)}, \quad (21)$$

respectively, having defined

$$\alpha_k^{(m,1)} = \sum_{c_k^{(m)} \in \mathcal{C}} c_{h,k}^{(m,1)} P_d^{(c)}(c_k^{(m)}) \quad (22)$$

$$\beta_k^{(m,1)} = \sum_{c_k^{(m)} \in \mathcal{C}} \left| c_{h,k}^{(m,1)} \right|^2 P_d^{(c)}(c_k^{(m)}), \text{ for } m = \{1, 2\}. \quad (23)$$

Therefore,

$$p_d^{(\theta)}(\boldsymbol{\Theta}_k)$$
$$\approx \mathcal{N}(r_k^{(1)}; \mathbb{E}\{r_k^{(1)}|\theta_{t,k}^{(1)}, \theta_{t,k}^{(2)}, \theta_{r,k}^{(1)}\}, \gamma_k^{(1)}) \quad (24)$$
$$\propto \exp\left\{ -\frac{\left| r_k^{(1)} - \alpha_k^{(1,1)} e^{j(\theta_{t,k}^{(1)}+\theta_{r,k}^{(1)})} - \alpha_k^{(2,1)} e^{j(\theta_{t,k}^{(2)}+\theta_{r,k}^{(1)})} \right|^2}{\gamma_k^{(1)}} \right\}$$

$$\propto \exp\left\{ \frac{2\Re}{\gamma_k^{(1)}} \left[r_k^{(1)} \alpha_k^{(1,1)*} e^{-j(\theta_{t,k}^{(1)}+\theta_{r,k}^{(1)})} + r_k^{(1)} \alpha_k^{(2,1)*} e^{-j(\theta_{t,k}^{(2)}+\theta_{r,k}^{(1)})} \right.\right.$$
$$\left.\left. - \left|\alpha_k^{(2,1)}\alpha_k^{(1,1)*}\right| e^{j(\angle \alpha_k^{(2,1)}\alpha_k^{(1,1)*}+\theta_{t,k}^{(2)}-\theta_{t,k}^{(1)})} \right] \right\}$$

$$= \exp\left\{ \frac{2\Re}{\gamma_k^{(1)}} \left[r_k^{(1)} \alpha_k^{(1,1)*} e^{-j(\theta_{t,k}^{(1)}+\theta_{r,k}^{(1)})} + r_k^{(1)} \alpha_k^{(2,1)*} e^{-j(\theta_{t,k}^{(2)}+\theta_{r,k}^{(1)})} \right.\right.$$
$$\left.\left. - \left|\alpha_k^{(2,1)}\alpha_k^{(1,1)*}\right| e^{j(\angle r_k^{(1)}\alpha_k^{(1,1)*}-\angle r_k^{(1)}\alpha_k^{(2,1)*}+\theta_{t,k}^{(2)}-\theta_{t,k}^{(1)})} \right] \right\} \quad (25)$$

$$\triangleq \exp\left\{ \Re\left[(x_k^{(1)} e^{-j\theta_{t,k}^{(1)}} + x_k^{(2)} e^{-j\theta_{t,k}^{(2)}}) e^{-j\theta_{r,k}^{(1)}} \right.\right.$$
$$\left.\left. - x_k^{(3)} e^{-j(\theta_{t,k}^{(1)}-\theta_{t,k}^{(2)})} \right] \right\} \quad (26)$$

$$= \exp\left\{ \Re\left[x_k^{(1)} e^{-j\theta_k^{(1,1)}} + x_k^{(2)} e^{-j\theta_k^{(2,1)}} \right.\right.$$
$$\left.\left. - x_k^{(3)} e^{-j(\theta_k^{(1,1)}-\theta_k^{(2,1)})} \right] \right\}. \quad (27)$$

In (25), we exploit that $\angle \alpha_k^{(2,1)}\alpha_k^{(1,1)*} = \angle r_k^{(1)}\alpha_k^{(1,1)*} - \angle r_k^{(1)}\alpha_k^{(2,1)*}$, and in (26) we define

$$x_k^{(1)} \triangleq \frac{2}{\gamma_k^{(1)}} \left| r_k^{(1)}\alpha_k^{(1,1)*} \right| e^{j\angle r_k^{(1)}\alpha_k^{(1,1)*}}$$

$$x_k^{(2)} \triangleq \frac{2}{\gamma_k^{(1)}} \left| r_k^{(1)}\alpha_k^{(2,1)*} \right| e^{j\angle r_k^{(1)}\alpha_k^{(2,1)*}}$$

$$x_k^{(3)} \triangleq \frac{2}{\gamma_k^{(1)}} \left| \alpha_k^{(2,1)}\alpha_k^{(1,1)*} \right| e^{j(\angle r_k^{(1)}\alpha_k^{(1,1)*}-\angle r_k^{(1)}\alpha_k^{(2,1)*})}. \quad (28)$$

In (27), the message $p_\text{d}^{(\theta)}(\boldsymbol{\Theta}_k)$ is re-expressed as the cosine variant of the unnormalized bivariate Tikhonov distribution of $\theta_k^{(1,1)}, \theta_k^{(2,1)}$ [25]. The distribution is completely characterized by $x_k^{(1)}, x_k^{(2)}, x_k^{(3)}$ and its parameters are the following. $\angle x_k^{(1)}$ and $1/|x_k^{(1)}|$ are the mean and variance of $\theta_k^{(1,1)}$, respectively, $\angle x_k^{(2)}$ and $1/|x_k^{(2)}|$ are the mean and variance of $\theta_k^{(2,1)}$, respectively, and $x_k^{(3)}$ is related to the correlation between $\theta_k^{(1,1)}$ and $\theta_k^{(2,1)}$, which can have an arbitrary magnitude and has to satisfy the constraint $\angle x_k^{(3)} = \angle x_k^{(1)} - \angle x_k^{(2)}$. The estimates of $\theta_k^{(1,1)}, \theta_k^{(2,1)}$ and their covariance mentioned are based on the received signal $r_k^{(1)}$. In our algorithm, we exploit the functional form of $p_\text{d}^{(\theta)}(\boldsymbol{\Theta}_k)$ in (26) to determine the other messages.

A. Forward Recursion

In the sequel, based on $p_\text{d}^{(\theta)}(\boldsymbol{\Theta}_k)$ and motivated by the choice of the cannonical pdf in [8], we determine the message $p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_k)$, which is constrained to be a bivariate Tikhonov pdf. Computation of the parameters of $p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_k)$ renders the estimates of $\theta_k^{(1,1)}, \theta_k^{(2,1)}$ and their covariance based on the received signals $[r_1^{(1)}, \dots, r_L^{(1)}]$ in the forward direction – this is referred to as the forward recursion. The message is evaluated as

$$p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_k) = \int_{\boldsymbol{\Theta}_{k-1}} p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_{k-1}) p_\text{d}^{(\theta)}(\boldsymbol{\Theta}_{k-1}) p_\Delta(\boldsymbol{\Theta}_k - \boldsymbol{\Theta}_{k-1}) \text{d}\boldsymbol{\Theta}_{k-1}.$$

(29)

Assume that $p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_{k-1})$ is the cosine variant of the bivariate Tikhonov distribution and is given as

$$p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_{k-1}) \propto \exp\Big\{\Re\Big[\big(a_{\text{f},k-1}^{(1,1)} e^{-\jmath\theta_{\text{t},k-1}^{(1)}} + a_{\text{f},k-1}^{(2,1)} e^{-\jmath\theta_{\text{t},k-1}^{(2)}}\big) \\ \cdot e^{-\jmath\theta_{\text{r},k-1}^{(1)}} - \tilde{a}_{\text{f},k-1}^{(1,2)} e^{-\jmath(\theta_{\text{t},k-1}^{(1)} - \theta_{\text{t},k-1}^{(2)})}\Big]\Big\},$$

(30)

where $\angle\tilde{a}_{\text{f},k-1}^{(1,2)} = \angle a_{\text{f},k-1}^{(1,1)} - \angle a_{\text{f},k-1}^{(2,1)}$. In (30), $\angle a_{\text{f},k-1}^{(1)}, \angle a_{\text{f},k-1}^{(2)}$ and $1/|a_{\text{f},k-1}^{(1)}|, 1/|a_{\text{f},k-1}^{(2)}|$ correspond to the predicted estimates and the variances of $\theta_{k-1}^{(1,1)}, \theta_{k-1}^{(2,1)}$, respectively, based on the received signals $[r_1^{(1)}, \dots, r_{k-1}^{(1)}]$, and $\Re\{\tilde{a}_{\text{f},k}^{(1,2)}\}$ is the predicted correlation between $\theta_{k-1}^{(1,1)}, \theta_{k-1}^{(2,1)}$.

From the bivariate Tikhonov pdf-based approximation enforced as $\angle\tilde{a}_{\text{f},k-1}^{(1,2)} = \angle a_{\text{f},k-1}^{(1,1)} - \angle a_{\text{f},k-1}^{(2,1)}$, we have that the correlation, $\Re\{\tilde{a}_{\text{f},k-1}^{(1,2)}\}$, changes with $\angle a_{\text{f},k-1}^{(1,1)} - \angle a_{\text{f},k-1}^{(2,1)}$. Intuitively, this makes sense for the problem setup that we have: consider that the phase noise at the transmitter is very small compared to the receiver phase noise, i.e., $\theta_{k-1}^{(2,1)} = \theta_{\text{t},k-1}^{(2)} + \theta_{\text{r},k-1}^{(1)} \approx \theta_{\text{r},k-1}^{(1)}$, $\theta_{k-1}^{(1,1)} = \theta_{\text{t},k-1}^{(1)} + \theta_{\text{r},k-1}^{(1)} \approx \theta_{\text{r},k-1}^{(1)}$. Then, $\angle a_{\text{f},k-1}^{(1,1)} - \angle a_{\text{f},k-1}^{(2,1)} \approx 0$, and the correlation between $\theta_{k-1}^{(1,1)}$ and $\theta_{k-1}^{(2,1)}$ is maximized. However, as the phase noise at the transmitter becomes more severe, $\angle a_{\text{f},k-1}^{(1,1)} - \angle a_{\text{f},k-1}^{(2,1)}$ becomes more divergent, and the correlation between $\theta_{k-1}^{(1,1)}$ and $\theta_{k-1}^{(2,1)}$ reduces.

Now compute the product $p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_{k-1}) p_\text{d}^{(\theta)}(\boldsymbol{\Theta}_{k-1})$ in (29) as

$$p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_{k-1}) p_\text{d}^{(\theta)}(\boldsymbol{\Theta}_{k-1}) \\ = \exp\Big\{\Re\Big[\big((x_{k-1}^{(1)} + a_{\text{f},k-1}^{(1,1)})e^{-\jmath\theta_{\text{t},k-1}^{(1)}} + (x_{k-1}^{(2)} + a_{\text{f},k-1}^{(2,1)})e^{-\jmath\theta_{\text{t},k-1}^{(2)}}\big) \\ \cdot e^{-\jmath\theta_{\text{r},k-1}^{(1)}} - (x_{k-1}^{(3)} + \tilde{a}_{\text{f},k-1}^{(1,2)})e^{-\jmath(\theta_{\text{t},k-1}^{(1)} - \theta_{\text{t},k-1}^{(2)})}\Big]\Big\} \\ \triangleq \exp\Big\{\Re\Big[(y_{k-1}^{(1)} e^{-\jmath\theta_{\text{t},k-1}^{(1)}} + y_{k-1}^{(2)} e^{-\jmath\theta_{\text{t},k-1}^{(2)}})e^{-\jmath\theta_{\text{r},k-1}^{(1)}} \\ - y_{k-1}^{(3)} e^{-\jmath(\theta_{\text{t},k-1}^{(1)} - \theta_{\text{t},k-1}^{(2)})}\Big]\Big\}.$$

(31)

In (31), $\angle y_{k-1}^{(1)}$, $\angle y_{k-1}^{(2)}$ and $1/|y_{k-1}^{(1)}|$, $1/|y_{k-1}^{(2)}|$ correspond to the predicted estimates and the variances of $\theta_{k-1}^{(1,1)}$, $\theta_{k-1}^{(2,1)}$, respectively, based on the received signals $[r_1^{(1)}, \dots, r_k^{(1)}]$, and $y_{k-1}^{(3)}$ gives a measure of the predicted correlation between $\theta_{k-1}^{(1,1)}$, $\theta_{k-1}^{(2,1)}$.

Note that bivariate Tikhonov distributions are not closed under the product operation [26], i.e., the product of two bivariate Tikhonov distributions is not another bivariate Tikhonov distribution. However in [26], it was conjectured that the product is approximately another bivariate Tikhonov pdf. The quality of this approximation cannot be assessed analytically. As we shall see in Section VIII, the performance of the SPA-based receiver that is developed in this section is better than the other algorithms available in literature, implying that the approximations made are meaningful. This pragmatic approach of motivating the quality of the approximations by performance analysis of the algorithm is a common and well-accepted method in the Bayesian filtering literature [27]–[29]

Based on the conjecture in [26], we approximate the product in (31) as a bivariate Tikhonov distribution, where we impose that $\angle y_{k-1}^{(3)} \approx \angle y_{k-1}^{(1)} - \angle y_{k-1}^{(2)}$, as required for a bivariate Tikhonov distribution. Using this assumption, we compute $p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_k)$ in (29) as

$$p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_k) \\ = \int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi} \exp\Big\{\Re\Big[(y_{k-1}^{(1)} e^{-\jmath\theta_{\text{t},k-1}^{(1)}} + y_{k-1}^{(2)} e^{-\jmath\theta_{\text{t},k-1}^{(2)}}) \\ \cdot e^{-\jmath\theta_{\text{r},k-1}^{(1)}} - y_{k-1}^{(3)} e^{-\jmath(\theta_{\text{t},k-1}^{(1)} - \theta_{\text{t},k-1}^{(2)})}\Big]\Big\} \\ \cdot p_\Delta(\theta_{\text{t},k}^{(1)} - \theta_{\text{t},k-1}^{(1)}) p_\Delta(\theta_{\text{t},k}^{(2)} - \theta_{\text{t},k-1}^{(2)}) \\ \cdot p_\Delta(\theta_{\text{r},k}^{(1)} - \theta_{\text{r},k-1}^{(1)}) \text{d}\theta_{\text{t},k-1}^{(1)} \text{d}\theta_{\text{t},k-1}^{(2)} \text{d}\theta_{\text{r},k-1}^{(1)}.$$

(32)

For the discrete Wiener phase noise process considered, we show in Appendix A that $p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_k)$ is approximately a bivariate Tikhonov distribution given by

$$p_\text{f}^{(\theta)}(\boldsymbol{\Theta}_k) \propto \exp\Big\{\Re\Big[(a_{\text{f},k}^{(1,1)} e^{-\jmath\theta_{\text{t},k}^{(1)}} + a_{\text{f},k}^{(2,1)} e^{-\jmath\theta_{\text{t},k}^{(2)}})e^{-\jmath\theta_{\text{r},k}^{(1)}} \\ - \tilde{a}_{\text{f},k}^{(1,2)} e^{-\jmath(\theta_{\text{t},k}^{(1)} - \theta_{\text{t},k}^{(2)})}\Big]\Big\},$$

(33)

where it is assumed that $\angle\tilde{a}_{\text{f},k}^{(1,2)} = \angle a_{\text{f},k}^{(1,1)} - \angle a_{\text{f},k}^{(2,1)}$. The parameters $a_{\text{f},k}^{(1,1)}, a_{\text{f},k}^{(2,1)}, \tilde{a}_{\text{f},k}^{(1,2)}$ are recursively updated in the

forward direction as

$$a_{\text{f},k}^{(m,1)} = \frac{\bar{\bar{a}}_{\text{f},k}^{(m,1)}}{1 + \sigma_{\text{t}}^2 \left| \left| \bar{\bar{a}}_{\text{f},k}^{(m,1)} \right| - \left| \tilde{\bar{a}}_{\text{f},k}^{(1,2)} \right| \right|}, \quad m \in \{1,2\}$$

$$\tilde{a}_{\text{f},k}^{(1,2)} = \frac{\tilde{\bar{a}}_{\text{f},k}^{(1,2)}}{\prod_{m=1}^{2} \left(1 + \sigma_{\text{t}}^2 \left| \left| \bar{\bar{a}}_{\text{f},k}^{(m,1)} \right| - \left| \tilde{\bar{a}}_{\text{f},k}^{(1,2)} \right| \right| \right)}, \tag{34}$$

where

$$\bar{\bar{a}}_{\text{f},k}^{(m,1)} = \frac{\bar{a}_{\text{f},k}^{(m,1)}}{1 + \sigma_{\text{r}}^2 \left| \left| \bar{a}_{\text{f},k}^{(1,1)} \right| + \left| \bar{a}_{\text{f},k}^{(2,1)} \right| \right|}.$$

$$\bar{a}_{\text{f},k}^{(m,1)} = a_{\text{f},k-1}^{(m,1)} + \frac{2}{\gamma_{k-1}^{(1)}} r_{k-1}^{(1)} \alpha_{k-1}^{(m,1)*}$$

$$\tilde{\bar{a}}_{\text{f},k}^{(1,2)} = \tilde{a}_{\text{f},k-1}^{(1,2)} + \frac{2}{\gamma_{k-1}^{(1)}} \alpha_{k-1}^{(2,1)} \alpha_{k-1}^{(1,1)*} \tag{35}$$

B. Backward Recursion

The parameters of $p_{\text{b}}^{(\theta)}(\boldsymbol{\Theta}_k)$ are computed based on the received signals $[r_L^{(1)}, \ldots, r_1^{(1)}]$ in the backward direction. The message $p_{\text{b}}^{(\theta)}(\boldsymbol{\Theta}_k)$ in (14) is evaluated as

$$p_{\text{b}}^{(\theta)}(\boldsymbol{\Theta}_k) \propto \exp\Big\{ \Re \Big[(a_{\text{b},k}^{(1,1)} e^{-\jmath\theta_{\text{t},k}^{(1)}} + a_{\text{b},k}^{(2,1)} e^{-\jmath\theta_{\text{t},k}^{(2)}}) e^{-\jmath\theta_{\text{r},k}^{(1)}}$$
$$- \tilde{a}_{\text{b},k}^{(1,2)} e^{-\jmath(\theta_{\text{t},k}^{(1)} - \theta_{\text{t},k}^{(2)})} \Big] \Big\}, \tag{36}$$

where it is assumed that $\angle \tilde{a}_{\text{b},k}^{(1,2)} = \angle a_{\text{b},k}^{(1,1)} - \angle a_{\text{b},k}^{(2,1)}$. The parameters of $p_{\text{b}}^{(\theta)}(\boldsymbol{\Theta}_k)$ are recursively updated in the backward direction as

$$a_{\text{b},k}^{(m,1)} = \frac{\bar{\bar{a}}_{\text{b},k}^{(m,1)}}{1 + \sigma_{\text{t}}^2 \left| \left| \bar{\bar{a}}_{\text{b},k}^{(m,1)} \right| - \left| \tilde{\bar{a}}_{\text{b},k}^{(1,2)} \right| \right|}, \quad m \in \{1,2\}$$

$$\tilde{a}_{\text{b},k}^{(1,2)} = \frac{\tilde{\bar{a}}_{\text{b},k}^{(1,2)}}{\prod_{m=1}^{2} \left(1 + \sigma_{\text{t}}^2 \left| \left| \bar{\bar{a}}_{\text{b},k}^{(m,1)} \right| - \left| \tilde{\bar{a}}_{\text{b},k}^{(1,2)} \right| \right| \right)}, \tag{37}$$

where

$$\bar{\bar{a}}_{\text{b},k}^{(m,1)} = \frac{\bar{a}_{\text{b},k}^{(m,1)}}{1 + \sigma_{\text{r}}^2 \left| \left| \bar{a}_{\text{b},k}^{(1,1)} \right| + \left| \bar{a}_{\text{b},k}^{(2,1)} \right| \right|}$$

$$\bar{a}_{\text{b},k}^{(m,1)} = a_{\text{b},k+1}^{(m,1)} + \frac{2}{\gamma_{k+1}^{(1)}} r_{k+1}^{(1)} \alpha_{k+1}^{(m,1)*}$$

$$\tilde{\bar{a}}_{\text{b},k}^{(1,2)} = \tilde{a}_{\text{b},k+1}^{(1,2)} + \frac{2}{\gamma_{k+1}^{(1)}} \alpha_{k+1}^{(2,1)} \alpha_{k+1}^{(1,1)*}. \tag{38}$$

C. Computation of $P_{\text{u}}^{(c)}(\mathbf{c}_k)$

Based on the messages $p_{\text{f}}^{(\theta)}(\boldsymbol{\Theta}_k)$ and $p_{\text{b}}^{(\theta)}(\boldsymbol{\Theta}_k)$, we compute $P_{\text{u}}^{(c)}(\mathbf{c}_k)$ in (15) as

$$P_{\text{u}}^{(c)}(\mathbf{c}_k) = \int_{\boldsymbol{\Theta}_k} p_{\text{f}}^{(\theta)}(\boldsymbol{\Theta}_k) p_{\text{b}}^{(\theta)}(\boldsymbol{\Theta}_k) p(\mathbf{r}_k|\mathbf{c}_k, \boldsymbol{\Theta}_k) d\boldsymbol{\Theta}_k$$

$$\propto \exp\left(-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0} \right)$$

$$\cdot \int_{\boldsymbol{\Theta}_k} \exp\Big\{ \Re \Big[(z_k^{(1,1)} e^{-\jmath\theta_{\text{t},k}^{(1)}} + z_k^{(2,1)} e^{-\jmath\theta_{\text{t},k}^{(2)}}) e^{-\jmath\theta_{\text{r},k}^{(1)}}$$
$$- \tilde{z}_k^{(1,2)} e^{-\jmath(\theta_{\text{t},k}^{(1)} - \theta_{\text{t},k}^{(2)})} \Big] \Big\} d\boldsymbol{\Theta}_k$$

$$\propto \exp\left(-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0} \right) \text{I}_0 \left(\left| z_k^{(1,1)} \right| + \left| z_k^{(2,1)} \right| \right)$$

$$\cdot \text{I}_0 \left(\left| \tilde{z}_k^{(1,2)} \right| \right), \tag{39}$$

where $\text{I}_0(\cdot)$ is the zeroth order modified Bessel function, and we define

$$z_k^{(m,1)} \triangleq a_{\text{f},k}^{(m,1)} + a_{\text{b},k}^{(m,1)} + \frac{2}{N_0} r_k^{(1)} c_{\text{h},k}^{(m,1)*}, \quad m \in \{1,2\}$$

$$\tilde{z}_k^{(1,2)} \triangleq \tilde{a}_{\text{f},k}^{(1,2)} + \tilde{a}_{\text{b},k}^{(1,2)} + \frac{2}{N_0} c_{\text{h},k}^{(2,1)} c_{\text{h},k}^{(1,1)*}. \tag{40}$$

D. Generalization to Arbitrary N_{t} and N_{r} values

Based on (34), we can generalize the forward recursions for the case of arbitrary N_{t} and N_{r} values as

$$a_{\text{f},k}^{(m,n)} = \frac{\bar{\bar{a}}_{\text{f},k}^{(m,n)}}{1 + \sigma_{\text{t}}^2 \left| \sum_{n=1}^{N_{\text{r}}} \left| \bar{\bar{a}}_{\text{f},k}^{(m,n)} \right| - \sum_{\substack{l=1 \\ l \neq m}}^{N_{\text{t}}} \left| \tilde{\bar{a}}_{\text{f},k}^{(m,l)} \right| \right|}$$

$$\tilde{a}_{\text{f},k}^{(m,l)} = \frac{\tilde{\bar{a}}_{\text{f},k}^{(m,l)}}{\prod_{m=1}^{N_{\text{t}}} \left(1 + \sigma_{\text{t}}^2 \left| \sum_{n=1}^{N_{\text{r}}} \left| \bar{\bar{a}}_{\text{f},k}^{(m,n)} \right| - \sum_{\substack{l=1 \\ l \neq m}}^{N_{\text{t}}} \left| \tilde{\bar{a}}_{\text{f},k}^{(m,l)} \right| \right| \right)}, \tag{41}$$

where $m, l \in \{1, \ldots, N_{\text{t}}\}$ and $n \in \{1, \ldots, N_{\text{r}}\}$ and

$$\bar{\bar{a}}_{\text{f},k}^{(m,n)} = \frac{\bar{a}_{\text{f},k}^{(m,n)}}{1 + \sigma_{\text{r}}^2 \sum_{m=1}^{N_{\text{t}}} \left| \bar{a}_{\text{f},k}^{(m,n)} \right|}$$

$$\bar{a}_{\text{f},k}^{(m,n)} = a_{\text{f},k-1}^{(m,n)} + \frac{2}{\gamma_{k-1}^{(n)}} r_{k-1}^{(n)} \alpha_{k-1}^{(m,n)*}$$

$$\tilde{\bar{a}}_{\text{f},k}^{(m,l)} = \tilde{a}_{\text{f},k-1}^{(m,l)} + \frac{2 N_{\text{r}}}{\gamma_{k-1}^{(n)}} \alpha_{k-1}^{(l,n)} \alpha_{k-1}^{(m,n)*}. \tag{42}$$

Similarly, we can generalize the backward recursion as

$$a_{\text{b},k}^{(m,n)} = \frac{\bar{\bar{a}}_{\text{b},k}^{(m,n)}}{1 + \sigma_{\text{t}}^2 \left| \sum_{n=1}^{N_{\text{r}}} \left| \bar{\bar{a}}_{\text{b},k}^{(m,n)} \right| - \sum_{\substack{l=1 \\ l \neq m}}^{N_{\text{t}}} \left| \tilde{\bar{a}}_{\text{b},k}^{(m,l)} \right| \right|}$$

$$\tilde{a}_{\text{b},k}^{(m,l)} = \frac{\tilde{\bar{a}}_{\text{b},k}^{(m,l)}}{\prod_{m=1}^{N_{\text{t}}} \left(1 + \sigma_{\text{t}}^2 \left| \sum_{n=1}^{N_{\text{r}}} \left| \bar{\bar{a}}_{\text{b},k}^{(m,n)} \right| - \sum_{\substack{l=1 \\ l \neq m}}^{N_{\text{t}}} \left| \tilde{\bar{a}}_{\text{b},k}^{(m,l)} \right| \right| \right)}, \tag{43}$$

where

$$\bar{\bar{a}}_{\mathrm{b},k}^{(m,n)} = \frac{\bar{a}_{\mathrm{b},k}^{(m,n)}}{1 + \sigma_{\mathrm{r}}^2 \sum\limits_{m=1}^{N_{\mathrm{t}}} \left| \bar{a}_{\mathrm{b},k}^{(m,n)} \right|}$$

$$\bar{a}_{\mathrm{b},k}^{(m,n)} = a_{\mathrm{b},k+1}^{(m,n)} + \frac{2}{\gamma_{k+1}^{(n)}} r_{k+1}^{(n)} \alpha_{k+1}^{(m,n)*}$$

$$\tilde{\bar{a}}_{\mathrm{b},k}^{(m,l)} = \tilde{a}_{\mathrm{f},k+1}^{(m,l)} + \frac{2N_{\mathrm{r}}}{\gamma_{k+1}^{(n)}} \alpha_{k+1}^{(l,n)} \alpha_{k+1}^{(m,n)*}. \tag{44}$$

Note that the MAP estimate of phase noise processes at each time instant k is given as $\arg\max_{\boldsymbol{\Theta}_k} \; p_{\mathrm{f}}^{(\theta)}(\boldsymbol{\Theta}_k) p_{\mathrm{b}}^{(\theta)}(\boldsymbol{\Theta}_k) p_{\mathrm{d}}^{(\theta)}(\boldsymbol{\Theta}_k)$, and is not used in the receiver algorithm explicitly. The generalization of $P_{\mathrm{u}}^{(c)}(\mathbf{c}_k)$ is given as

$$P_{\mathrm{u}}^{(c)}(\mathbf{c}_k) \propto \exp\left(-N_{\mathrm{r}} \sum_{m=1}^{N_{\mathrm{t}}} \frac{|c_{\mathrm{h},k}^{(m,n)}|^2}{N_0} \right) \prod_{n=1}^{N_{\mathrm{r}}} I_0 \left(\sum_{m=1}^{N_{\mathrm{t}}} \left| z_k^{(m,n)} \right| \right)$$
$$\prod_{\substack{m=1, \\ l>m}}^{N_{\mathrm{t}}} I_0 \left(\left| \tilde{z}_k^{(m,l)} \right| \right), \tag{45}$$

where

$$z_k^{(m,n)} \triangleq a_{\mathrm{f},k}^{(m,n)} + a_{\mathrm{b},k}^{(m,n)} + \sum_{n=1}^{N_{\mathrm{r}}} \frac{2}{N_0} r_k^{(n)} c_{\mathrm{h},k}^{(m,n)*}$$

$$\tilde{z}_k^{(m,l)} \triangleq \tilde{a}_{\mathrm{f},k}^{(m,l)} + \tilde{a}_{\mathrm{b},k}^{(m,l)} + \sum_{n=1}^{N_{\mathrm{r}}} \frac{2}{N_0} c_{\mathrm{h},k}^{(l,n)} c_{\mathrm{h},k}^{(m,n)*}. \tag{46}$$

For future reference, we refer to the SPA-based algorithm for approximate MAP detection as SPA-MAP. This algorithm is a generalization of the algorithm presented for single-antenna systems in [8]. We summarize one iteration of the SPA-MAP algorithm as follows.

Step 1) Evaluate the coefficients $\alpha_k^{(m,n)}, \beta_k^{(m,n)}$ in (22) using the a priori probabilities $P_{\mathrm{d}}^{(c)}(\mathbf{c}_k)$ of the transmitted symbols $c_k^{(i)} \forall k \in \{1, \ldots, L\}$ and $m \in \{1, \ldots, N_{\mathrm{t}}\}, n \in \{1, \ldots, N_{\mathrm{r}}\}$.
Step 2) Recursively update the parameters in (41) in the forward direction using the received signals $[r_1^{(n)}, \ldots, r_L^{(n)}] \forall n \in \{1, \ldots, N_{\mathrm{r}}\}$.
Step 3) Recursively update the parameters in (43) in the backward direction using the received signals $[r_L^{(n)}, \ldots, r_1^{(n)}] \forall n \in \{1, \ldots, N_{\mathrm{r}}\}$.
Step 4) Evaluate $P_{\mathrm{u}}^{(c)}(\mathbf{c}_k)$ in (45) $\forall k \in \{1, \ldots, L\}$.
Step 5) Update $P_{\mathrm{d}}^{(c)}(\mathbf{c}_k)$ using $P_{\mathrm{u}}^{(c)}(\mathbf{c}_k)$, if a soft-input soft-output decoder is used $\forall k \in \{1, \ldots, L\}$.

V. APPROXIMATE MAP DETECTION BASED ON THE SMOOTHER-DETECTOR STRUCTURE

In this section, we present a receiver algorithm that uses the smoother-detector structure, as in [10], to approximate the MAP symbol detector. As required by this structure, a smoother like the EKS is used to track the discrete Wiener phase noise process. Then the a posteriori pdf from the smoother is used for deriving the approximate MAP detector in (8). Specifically, let $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}})$ denote the a posteriori phase noise pdf provided by the smoother. We approximate $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}})$ by constraining it to a specific family of pdfs that renders the integral in (8) tractable.

We first consider the case where $N_{\mathrm{t}} = 2$ and $N_{\mathrm{r}} = 1$, and, with a slight abuse of notation, we let $\boldsymbol{\Theta}_k = [\theta_k^{(1,1)}, \theta_k^{(2,1)}]$ and $\mathbf{c}_k = [c_k^{(1)}, c_k^{(2)}]$. The pdf $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}})$ is modeled as a bivariate Gaussian pdf, i.e., $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}}) = \mathcal{N}(\boldsymbol{\Theta}_k; \widehat{\boldsymbol{\Theta}}_k, \mathbf{P}_{k,n})$, where $\hat{\theta}_k^{(m,n)} \in \widehat{\boldsymbol{\Theta}}_k$ and $P_{k,n}^{(m,l)} \in \mathbf{P}_k$, with

$$\hat{\theta}_k^{(m,n)} = \mathbb{E}_{p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}})}\{\theta_k^{(m,n)}\},$$
$$P_{k,n}^{(m,l)} = \mathbb{E}_{p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}})}\{(\theta_k^{(m,n)} - \hat{\theta}_k^{(m,n)})(\theta_k^{(l,n)} - \hat{\theta}_k^{(l,n)})\}.$$
$$\text{for } n = 1 \text{ and } m, l \in \{1, 2\} \tag{47}$$

Thus, the approximate MAP detector is written as

$$\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k \in \mathcal{C}} \int\limits_{\boldsymbol{\Theta}_k} p(\mathbf{r}_k|\mathbf{c}_k, \boldsymbol{\Theta}_k) p(\boldsymbol{\Theta}_k|\mathbf{c}_k, \bar{\mathbf{r}}_k) \mathrm{d}\boldsymbol{\Theta}_k, \tag{48}$$

$$\approx \arg\max_{\mathbf{c}_k \in \mathcal{C}} \int\limits_{\boldsymbol{\Theta}_k} p(\mathbf{r}_k|\mathbf{c}_k, \boldsymbol{\Theta}_k) p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}}) \mathrm{d}\boldsymbol{\Theta}_k \tag{49}$$

$$= \arg\max_{\mathbf{c}_k \in \mathcal{C}} \int\limits_{\boldsymbol{\Theta}_k} p(\mathbf{r}_k|\mathbf{c}_k, \boldsymbol{\Theta}_k) \mathcal{N}(\boldsymbol{\Theta}_k; \widehat{\boldsymbol{\Theta}}_k, \hat{\mathbf{P}}_k) \mathrm{d}\boldsymbol{\Theta}_k \tag{50}$$

$$\approx \arg\max_{\mathbf{c}_k \in \mathcal{C}} \; \exp\left\{ -\frac{|c_{\mathrm{h},k}^{(1,1)}|^2 + |c_{\mathrm{h},k}^{(2,1)}|^2}{N_0} \right\}$$
$$\cdot I_0 \left(|u_k^{(1,1)}| + |u_k^{(2,1)}| - |\tilde{u}_k^{(1,2)}| \right). \tag{51}$$

We refer the reader to Appendix B for the derivation of the result in (51). In (50), we define

$$u_k^{(m,1)} \triangleq \frac{2}{N_0} r_k^{(1)} c_{\mathrm{h},k}^{(m,n)*} + \frac{e^{j\hat{\theta}_k^{(m,1)}}}{P_{k,1}^{(m,1)}}, \; m \in \{1, 2\}$$

$$\tilde{u}_k^{(1,2)} \triangleq \frac{2}{N_0} c_{\mathrm{h},k}^{(2,1)} c_{\mathrm{h},k}^{(1,1)*} + \left| \tilde{\tilde{u}}_k^{(1,2)} \right| e^{j(\hat{\theta}_k^{(1,1)} - \hat{\theta}_k^{(2,1)})}, \tag{52}$$

where $\left| \tilde{\tilde{u}}_k^{(1,2)} \right|$ is obtained by solving [25]

$$P_{k,1}^{1,2} = \frac{-|\tilde{\tilde{u}}_k^{(1,2)}|}{\sqrt{\left(P_{k,1}^{(1,1)^{-1}} - |\tilde{u}_k^{(1,2)}| \right) \left(P_{k,1}^{(2,2)^{-1}} - |\tilde{u}_k^{(1,2)}| \right)}}. \tag{53}$$

The generalization of the approximate MAP detector in (51) to arbitrary N_{t} and N_{r} values is given as

$$\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k} \; \exp\left(-N_{\mathrm{r}} \sum_{m=1}^{N_{\mathrm{t}}} \frac{|c_{\mathrm{h},k}^{(m,n)}|^2}{N_0} \right)$$
$$\prod_{n=1}^{N_{\mathrm{r}}} I_0 \left(\sum_{m=1}^{N_{\mathrm{t}}} |u_k^{(m,n)}| - \sum_{\substack{m=1, \\ l>m}}^{N_{\mathrm{t}}} |\tilde{u}_k^{(m,l)}| \right). \tag{54}$$

For future reference, we refer to the smoother-detector algorithm for approximate MAP detection in (51) based on the Gaussian pdf assumption for phase noise as Gauss-MAP. When the a posteriori pdf $p(\boldsymbol{\Theta}_k|\bar{\mathbf{r}}_k)$ is considered to be a Dirac Delta function, $\delta(\boldsymbol{\Theta}_k - \widehat{\boldsymbol{\Theta}}_k)$, then the symbol detector in (50) reduces to the Euclidean distance-based detector that treats the phase noise estimate as the true value of the phase noise at the kth time instant, i.e.,

$$\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k \in \mathcal{C}} \int\limits_{\boldsymbol{\Theta}_k} p(\mathbf{r}_k|\mathbf{c}_k, \boldsymbol{\Theta}_k) \delta(\boldsymbol{\Theta}_k - \widehat{\boldsymbol{\Theta}}_k) \mathrm{d}\boldsymbol{\Theta}_k$$

$$= \arg\max_{\mathbf{c}_k \in \mathcal{C}} p(\mathbf{r}_k|\mathbf{c}_k, \widehat{\boldsymbol{\Theta}}_k). \tag{55}$$

In the sequel, we refer to the symbol-by-symbol detector comprising the smoother and the Euclidean distance-based detector in (55) as EUC-MAP.

For implementing the Gauss-MAP and the EUC-MAP, the EKS [32] is used to compute the estimate of the phase noise in each link and its respective variance as in (47). It is known from [13], that in a MIMO setup, the EKS achieves the Cramér-Rao lower bound (CRLB) in the data-aided estimation setup. In order to determine the linearized state space model for deriving the EKS, we model the symbol transmitted by the mth transmit antenna in the kth time instant as

$$c_k^{(m)} = \underline{c}_k^{(m)} + w_{c,k}^{(m)}, \tag{56}$$

where $\underline{c}_k^{(m)}$ is the symbol average from the detector and $w_{c,k}^{(m)}$ is the error associated with it, which is assumed to be Gaussian distributed, i.e., $w_{c,k}^{(m)} \sim \mathcal{N}(0, \sigma_c^2)$ [33], [34]. Define an $N_r \times N_t N_r$ matrix $\underline{\mathbf{C}}_k$ as

$$\underline{\mathbf{C}}_k = \begin{bmatrix} \underline{\mathbf{c}}_k & \mathbf{0}_{1\times N_t} & \cdots & \mathbf{0}_{1\times N_t} \\ \mathbf{0}_{1\times N_t} & \underline{\mathbf{c}}_k & \cdots & \mathbf{0}_{1\times N_t} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{1\times N_t} & \mathbf{0}_{1\times N_t} & \cdots & \underline{\mathbf{c}}_k \end{bmatrix}, \tag{57}$$

$$\underline{\mathbf{c}}_k = [\underline{c}_k^{(1)} \ldots \underline{c}_k^{(N_t)}]. \tag{58}$$

Thus the observation and the state space model considering the received signal at the nth receive antenna and the kth time instant for N_t transmit and N_r receive antennas is given as

$$\begin{aligned} r_k^{(n)} &= \sum_{m=1}^{N_t} e^{j\theta_k^{(m,n)}}(\underline{c}_k^{(m)} + w_{c,k}^{(m)})h^{(m,n)} + w_k^{(n)} \\ &\approx \sum_{m=1}^{N_t} e^{j\hat{\theta}_k^{(m,n)}}(1 + j(\theta_k^{(m,n)} - \hat{\theta}_k^{(m,n)}))\underline{c}_{h,k}^{(m,n)} + \tilde{w}_k^{(n)} \\ \theta_k^{(m,n)} &= \theta_{k-1}^{(m,n)} + \Delta_{t,k}^{(m)} + \Delta_{r,k}^{(n)}, \\ &\triangleq \theta_{k-1}^{(m,n)} + \Delta_k^{(m,n)}, \end{aligned} \tag{59}$$

where $\underline{c}_{h,k}^{(m,n)} \triangleq \underline{c}_k^{(m)} h^{(m,n)}$, $\tilde{w}_k^{(n)} \sim \mathcal{N}(0, N_0 + \sigma_c^2)$. The unknown state vector is $\boldsymbol{\Theta}_k$ and its estimate is $\widehat{\boldsymbol{\Theta}}_k$, the state noise vector is $[\Delta_k^{(1,1)}, \ldots, \Delta_k^{(N_t,1)}, \ldots, \Delta_k^{(N_t,N_r)}]$, all of which are of size $N_t N_r \times 1$. The state noise covariance matrix, denoted as $\mathbf{Q}$, is an $N_t N_r \times N_t N_r$ matrix, which is written as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{1,1} & \cdots & \mathbf{Q}_{1,N_r} \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_{N_r,1} & \cdots & \mathbf{Q}_{N_r,N_r} \end{bmatrix},$$

$$\mathbf{Q}_{a,b} = \begin{cases} \sigma_r^2 \mathbf{1}_{N_t\times N_t} + \sigma_t^2 \mathbf{I}_{N_t\times N_t}, & \text{if } a = b \\ \sigma_t^2 \mathbf{I}_{N_t\times N_t}, & \text{otherwise.} \end{cases}$$

Let $\mathbf{W} = \sigma_t^2 \mathbf{I}_{N_r\times N_r}$. Define the Jacobian matrix of size $N_r \times N_t N_r$ for the observation model in (59) as

$$\mathbf{H}(\widehat{\boldsymbol{\Theta}}_k) = \begin{bmatrix} \mathbf{h}_1 & \mathbf{0}_{1\times N_t} & \cdots & \mathbf{0}_{1\times N_t} \\ \mathbf{0}_{1\times N_t} & \mathbf{h}_2 & \cdots & \mathbf{0}_{1\times N_t} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{1\times N_t} & \mathbf{0}_{1\times N_t} & \cdots & \mathbf{h}_{N_r} \end{bmatrix}, \tag{60}$$

$$\mathbf{h}_n = [j\underline{c}_{h,k}^{(1,n)} e^{j\hat{\theta}_k^{(1,n)}} \cdots j\underline{c}_{h,k}^{(N_t,n)} e^{j\hat{\theta}_k^{(N_t,n)}}]. \tag{61}$$

Then the EKS equations are given in (62)-(67). $\mathbf{K}_k$ is the $N_t N_r \times N_r$ Kalman gain in the kth time instant. In the forward recursion, $\widehat{\boldsymbol{\Theta}}_k^f$ and $\mathbf{P}_k^f$ denote the estimate of the mean and covariance of $\boldsymbol{\Theta}_k$, respectively, based on the received signals $[\mathbf{r}_1, \ldots, \mathbf{r}_L]$. These are used in the backward recursion to obtain the smoothed estimate of the mean and covariance of $\boldsymbol{\Theta}_k$, denoted as $\widehat{\boldsymbol{\Theta}}_k$ and $\mathbf{P}_k$, respectively.

VI. VB Framework-Based Algorithm for Approximate MAP Detection

In this section, we develop a receiver algorithm for approximating the MAP symbol detector based on the VB framework as in [9] (see [30] for a nice tutorial on the VB framework). We consider arbitrary N_t and N_r, and with a slight abuse of notation, we define $\boldsymbol{\Theta}_k = [\theta_k^{(1,1)}, \ldots, \theta_k^{(m,n)}, \ldots, \theta_k^{(N_t,N_r)}]$ and $\mathbf{c}_k = [c_k^{(1)}, \ldots, c_k^{(N_t)}]$. Based on this framework, we first compute the log likelihood of $\bar{\mathbf{r}}$ as

$$\begin{aligned} \log p(\bar{\mathbf{r}}) &= \log \sum_{\bar{\mathbf{c}}} \int_{\bar{\boldsymbol{\Theta}}} p(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}, \bar{\mathbf{r}}) d\bar{\boldsymbol{\Theta}} \\ &= \log \sum_{\bar{\mathbf{c}}} \int_{\bar{\boldsymbol{\Theta}}} Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}) \frac{p(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}, \bar{\mathbf{r}})}{Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}})} d\bar{\boldsymbol{\Theta}} \\ &\overset{(a)}{\geq} \sum_{\bar{\mathbf{c}}} \int_{\bar{\boldsymbol{\Theta}}} Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}) \log \frac{p(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}, \bar{\mathbf{r}})}{Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}})} d\bar{\boldsymbol{\Theta}}. \end{aligned} \tag{68}$$

In (68), the Jensen's inequality is applied to lower bound the log likelihood; so when $Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}})$ is set to $P(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}|\bar{\mathbf{r}})$, the lower bound is achieved. Thus, our objective is to search over the various pdfs $Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}})$ can assume, such that the bound in (68) is as tight as possible. This procedure helps to approximate the MAP detector presented in (4) and (5). In order to reduce the search space, as in [9], [30], we constrain $Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}})$ to a family of factorized pdfs, i.e., we assume that $Q(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}) = q_{\mathbf{c}}(\bar{\mathbf{c}})q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}})$. This also corresponds to the assumption that $\bar{\mathbf{c}}$ and $\bar{\boldsymbol{\Theta}}$ are independent of each other given $\bar{\mathbf{r}}$. On one hand, some structure of the original problem is lost with this assumption. On the other hand, the optimization space is tremendously simplified thereby enabling to search for a pdf that would minimize the KL divergence measure with respect to $P(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}|\bar{\mathbf{r}})$. Note that it is difficult to provide a rigorous mathematical analysis on the quality of this approximation. However, the VB framework has been used in many prior works in order to design efficient receivers [9], [31].

The lower bound in (68) is rewritten as

$$\begin{aligned} \log p(\bar{\mathbf{r}}) &\geq \sum_{\bar{\mathbf{c}}} \int_{\bar{\boldsymbol{\Theta}}} q_{\mathbf{c}}(\bar{\mathbf{c}}) q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}}) \log \frac{p(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}, \bar{\mathbf{r}})}{q_{\mathbf{c}}(\bar{\mathbf{c}}) q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}})} d\bar{\boldsymbol{\Theta}} \\ &\triangleq \mathcal{H}(q_{\mathbf{c}}(\bar{\mathbf{c}}), q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}}), \bar{\mathbf{r}}), \end{aligned} \tag{69}$$

where $\mathcal{H}(q_{\mathbf{c}}(\bar{\mathbf{c}}), q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}}), \bar{\mathbf{r}})$ is referred to as the variational free energy – its maximization results in the minimization of the KL divergence between $q_{\mathbf{c}}(\bar{\mathbf{c}})q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}})$ and $p(\bar{\mathbf{c}}, \bar{\boldsymbol{\Theta}}|\bar{\mathbf{r}})$. To determine the factorized pdf, $q_{\mathbf{c}}(\bar{\mathbf{c}})$ and $q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}})$, that maximize $\mathcal{H}$, a coordinate ascent routine is used that maximizes with respect to one pdf while keeping the other fixed, in an alternating manner. Based on the functional derivatives of $\mathcal{H}$ with respect to the factorized pdf [9], the coordinate ascent routine involves the iterative

$$\mathbf{K}_k = (\mathbf{P}_{k-1}^{\mathrm{f}} + \mathbf{Q})\mathbf{H}^*(\widehat{\boldsymbol{\Theta}}_{k-1}^{\mathrm{f}})\left(\mathbf{H}(\widehat{\boldsymbol{\Theta}}_{k-1}^{\mathrm{f}})(\mathbf{P}_{k-1}^{\mathrm{f}} + \mathbf{Q})\mathbf{H}^*(\widehat{\boldsymbol{\Theta}}_{k-1}^{\mathrm{f}}) + \mathbf{W}\right)^{-1} \tag{62}$$

$$\widehat{\boldsymbol{\Theta}}_k^{\mathrm{f}} = \widehat{\boldsymbol{\Theta}}_{k-1}^{\mathrm{f}} + \Re\left\{\mathbf{K}_k\left[\mathbf{r}_{k-1} - \underline{\mathbf{C}}_k e^{\jmath\widehat{\boldsymbol{\Theta}}_{k-1}^{\mathrm{f}}}\right]\right\} \tag{63}$$

$$\mathbf{P}_k^{\mathrm{f}} = \Re\left\{(\mathbf{P}_{k-1}^{\mathrm{f}} + \mathbf{Q}) - \mathbf{K}_k\mathbf{H}^*(\widehat{\boldsymbol{\Theta}}_{k-1}^{\mathrm{f}})(\mathbf{P}_{k-1}^{\mathrm{f}} + \mathbf{Q})\right\} \tag{64}$$

$$\mathbf{J}_k = \mathbf{P}_k^{\mathrm{f}}\mathbf{I}_{N_{\mathrm{r}}N_{\mathrm{t}} \times N_{\mathrm{r}}N_{\mathrm{t}}}(\mathbf{P}_k^{\mathrm{f}} + \mathbf{Q})^{-1} \tag{65}$$

$$\widehat{\boldsymbol{\Theta}}_k = \widehat{\boldsymbol{\Theta}}_k^{\mathrm{f}} + \mathbf{J}_k(\widehat{\boldsymbol{\Theta}}_{k+1} - \widehat{\boldsymbol{\Theta}}_k^{\mathrm{f}}) \tag{66}$$

$$\mathbf{P}_k = \mathbf{P}_k^{\mathrm{f}} + \mathbf{J}_k(\mathbf{P}_{k+1} - (\mathbf{P}_k + \mathbf{Q}))\mathbf{J}_k^* \tag{67}$$

computation of

$$q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}}) \propto p(\bar{\boldsymbol{\Theta}})e^{\sum_{\bar{\mathbf{c}}} q_{\mathbf{c}}(\bar{\mathbf{c}}) \log P(\bar{\mathbf{r}}|\bar{\mathbf{c}},\bar{\boldsymbol{\Theta}})}$$
$$q_{\mathbf{c}}(\bar{\mathbf{c}}) \propto P(\bar{\mathbf{c}})e^{\int_{\bar{\boldsymbol{\Theta}}} q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}}) \log P(\bar{\mathbf{r}}|\bar{\mathbf{c}},\bar{\boldsymbol{\Theta}})\mathrm{d}\bar{\boldsymbol{\Theta}}}. \tag{70}$$

The coordinate ascent routine is ensured to converge to a fixed point [30], but global optimality is not guaranteed.

We can immediately see that the coordinate ascent routine results in a receiver algorithm that iteratively computes the a posteriori phase noise pdf and the symbol pmf as given in (70). To derive their respective functional forms, we consider the received signal model in (1). Based on (70), the factorized pdf of $\bar{\boldsymbol{\Theta}}$ is derived as

$$q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}}) \approx p(\bar{\boldsymbol{\Theta}}|\bar{\mathbf{r}},\underline{\mathbf{c}}), \text{ where,}$$
$$\mathbb{E}_{q_{\mathbf{c}}}\{\mathbf{c}_k\} = \underline{\mathbf{c}}_k, \ \underline{\mathbf{c}} = [\underline{\mathbf{c}}_1,\dots,\underline{\mathbf{c}}_L], \mathrm{Var}_{q_{\mathbf{c}}}\{\mathbf{c}_k\} \approx 0, \tag{71}$$

where $\underline{\mathbf{c}}$ denotes the sequence of symbol averages transmitted by all transmit antennas and is used for computing the factorized pdf $q_{\boldsymbol{\theta}}(\bar{\boldsymbol{\Theta}})$. Furthermore, these symbol averages are treated as the true transmitted symbols as imposed by the variance constraint. We refer the reader to Appendix C for the proof of this result.

The factorized pmf of $\bar{\mathbf{c}}$ is given by

$$q_{\mathbf{c}}(\bar{\mathbf{c}}) = C_{\boldsymbol{c}} \prod_{k=1}^{L} P(\mathbf{c}_k) \exp\{C_{\mathrm{temp}}^{(1)}\}$$

$$C_{\mathrm{temp}}^{(1)} = -\frac{1}{N_0}\left\{\sum_{n=1}^{N_{\mathrm{r}}}\left|r_k^{(n)} - \sum_{m=1}^{N_{\mathrm{t}}}c_{\mathrm{h},k}^{(m,n)}e^{\jmath\hat{\theta}_k^{(m,n)}}\right|^2 - \sum_{m=1}^{N_{\mathrm{t}}}\left|c_{\mathrm{h},k}^{(m,n)}\right|^2\right.$$

$$\left. \cdot P_{k,n}^{(m,m)} - \sum_{m=1}^{N_{\mathrm{t}}}\sum_{\substack{l=1\\l\neq m}}^{N_{\mathrm{t}}} c_{\mathrm{h},k}^{(m,n)}c_{\mathrm{h},k}^{(l,n)*}P_{k,n}^{(m,l)}e^{\jmath(\hat{\theta}_k^{(m,n)} - \jmath\hat{\theta}_k^{(l,n)})}\right\}. \tag{72}$$

In (72), we assume that the a priori symbol sequence probability factorizes fully, under the assumption that the symbols are independent as in an uncoded transmission. The constant $C_{\boldsymbol{c}}$ normalizes the pmf and is independent of the transmitted symbols. The estimates of the phase noise in each link and the covariance matrix $\mathbf{P}_{k,n}$ are obtained by using an EKS [32].

Thus, the approximate MAP detector based on the VB framework is

$$\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k} P(\mathbf{c}_k)\exp\{C_{\mathrm{temp}}^{(1)}\} \tag{73}$$

where $q_{\mathbf{c}}(\bar{\mathbf{c}})$ corresponds to the symbol pmf to which the coordinate ascent routine in (70) converges. For future reference, we refer to the approximate MAP detector that is derived based on the VB framework as VB-MAP.
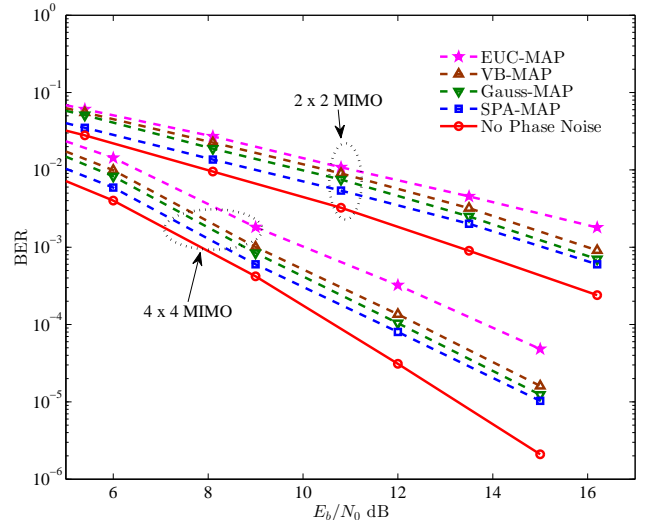

Fig. 2: BER performance for uncoded data transmission in different MIMO systems using the different receiver algorithms for $\sigma_{\mathrm{t}} = \sigma_{\mathrm{r}} = 4°$.

VII. COMPLEXITY ANALYSIS

In this section, we analyze the computational complexity of the proposed algorithms, namely the SPA-MAP (39), Gauss-MAP (51) and VB-MAP (72). The computational complexity of the algorithms is evaluated as the total number of additions and multiplications required to form the a posteriori symbol pmf, i.e., (45), (54) and (72), for all possible combination of the transmit symbols, plus the computations involved in the decoder (in the presence of a channel code).

Let the number of iterations between the decoder and the detector be denoted as L_{Det}, and let the number of iterations within the LDPC decoder be denoted as L_{Dec}. Furthermore, let the total number of addition and multiplication operations involved in the mapping and demapping functions and in the computation of the soft symbol for the detector be denoted as $\kappa_{\mathrm{map}}^{(\mathrm{T})}$. Consider that an LDPC code is used for coded transmission, and let $\kappa_{\mathrm{dec}}^{(\mathrm{T})}$ represent the total number of operations involved in one iteration within the LDPC decoder. Assume that the exponential function, the Bessel function, and the square root operations are implemented by means of a lookup table. For the SPA-MAP, denote the number of multiplications and additions involved as $\kappa_{\mathrm{spa}}^{(\mathrm{M})}$ and $\kappa_{\mathrm{spa}}^{(\mathrm{A})}$, respectively. Then, the total computational complexity for the SPA-based receiver is written as [36], [37] $\kappa_{\mathrm{spa}}^{(\mathrm{T})} = L_{\mathrm{Det}}\left(\kappa_{\mathrm{spa}}^{(\mathrm{M})} + \kappa_{\mathrm{spa}}^{(\mathrm{A})} + L_{\mathrm{Dec}}\kappa_{\mathrm{dec}}^{(\mathrm{T})} + \kappa_{\mathrm{map}}^{(\mathrm{T})}\right),$

where[1]

$$\kappa_{\text{spa}}^{(M)} = \underbrace{2N_t^3 + 2N_r^2 N_t + 8N_t^2 + 12N_t N_r}_{(41),(42),(43),(44)}$$

$$+ M^{N_t}\left(\underbrace{2N_r^2 N_t + 2N_t N_r^2 + N_t + N_r + 1}_{(45),(46)}\right) \quad (74)$$

$$\kappa_{\text{spa}}^{(A)} = \underbrace{8N_t^2 + 4N_r N_t + 2N_r N_t^2 + 2N_t^3 + 4N_t^2 N_r + 2N_r^2 N_t}_{(41),(42),(43),(44)}$$

$$+ M^{N_t}\left(\underbrace{2N_r^2 N_t + 2N_r N_t^2 + N_r N_t + 2N_t}_{(45),(46)}\right) \quad (75)$$

$$\kappa_{\text{map}}^{(T)} = M^{N_t}\left(\frac{M}{2}\log_2 M + \frac{M}{2} - 1 + N_t\right) \quad (76)$$

$$\kappa_{\text{dec}}^{(T)} = (2w_c + 3w_r - 3w_r r_c + 1)\log_2 M^{N_t}, \quad (77)$$

where w_c, w_r are the column and row weights of the LDPC parity check matrix, and r_c is the rate of the LDPC code.

The total computational complexity for the Gauss-MAP receiver is written as $\kappa_{\text{gauss}}^{(T)} = L_{\text{Det}}\left(\kappa_{\text{gauss}}^{(M)} + \kappa_{\text{gauss}}^{(A)} + L_{\text{Dec}}\kappa_{\text{dec}}^{(T)} + \kappa_{\text{map}}^{(T)}\right)$, where the number of multiplication operations $\kappa_{\text{gauss}}^{(M)}$ and addition operations $\kappa_{\text{gauss}}^{(A)}$ are determined as

$$\kappa_{\text{gauss}}^{(M)} = \underbrace{2N_t N_r}_{(60)} + \underbrace{2N_r^2 N_t(N_t N_r + N_r)}_{(62)} + N_r^3$$

$$+ \underbrace{2N_t N_r^2}_{(63)} + \underbrace{N_r^3 N_t^2(1 + N_t)}_{(64)} + \underbrace{3N_r^3 N_t^3}_{(65)}$$

$$+ \underbrace{N_r^2 N_t^2}_{(66)} + \underbrace{2N_r^3 N_t^3}_{(67)} + M^{N_t}\left(\underbrace{N_r + 4N_t^2 + 4N_r N_t}_{(54)}\right)$$
$$(78)$$

$$\kappa_{\text{gauss}}^{(A)} = \underbrace{N_t^2 N_r^2 + N_r(N_r N_t - 1)(2N_r N_t + N_r)}_{(62)}$$

$$+ \underbrace{N_r^2(1 + N_t(N_r - 1)) + N_r^3}_{(62)}$$

$$+ \underbrace{4N_r(N_t N_r - 1) + 2N_r + N_r N_t}_{(63)}$$

$$+ \underbrace{N_r^2 N_t^2(N_r(N_t - 1) - 1)}_{(64)}$$

$$+ \underbrace{N_r^3 N_t^3 + N_r^2 N_t^2 + 2N_r^2 N_t^2(N_r N_t - 1)}_{(65)}$$

$$+ \underbrace{2N_r N_t + N_r N_t(N_r N_t - 1)}_{(66)}$$

$$+ \underbrace{2N_r^2 N_t^2 + 2N_r^2 N_t^2(N_r N_t - 1)}_{(67)}$$

$$+ M^{N_t}\left(\underbrace{2N_t^2 + 2N_r N_t + N_t - 2N_r - 1}_{(54)}\right) \quad (79)$$

[1]Each term indicates the equations from which the computations arise.

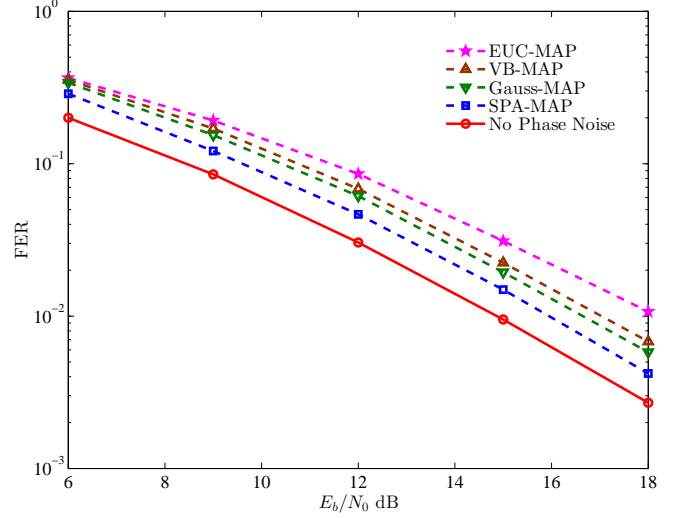Fig. 3: FER performance of a 2×1 system using different receiver algorithms for code rate $= 1/2$, $\sigma_t = \sigma_r = 4°$.

The computational complexity due to the EKS in the VB-MAP algorithm is the same as that in Gauss-MAP. Thus the total computational complexity for the VB-MAP receiver is $\kappa_{\text{vb}}^{(T)} = L_{\text{Det}}\left(\kappa_v^{(M)} + \kappa_{\text{vb}}^{(A)} + L_{\text{Dec}}\kappa_{\text{dec}}^{(T)} + \kappa_{\text{map}}^{(T)}\right)$, where

$$\kappa_{\text{vb}}^{(M)} = \underbrace{2N_t N_r}_{(60)} + \underbrace{2N_r^2 N_t(N_t N_r + N_r)}_{(62)} + N_r^3$$

$$+ \underbrace{2N_t N_r^2}_{(63)} + \underbrace{N_r^3 N_t^2(1 + N_t)}_{(64)} + \underbrace{3N_r^3 N_t^3}_{(65)}$$

$$+ \underbrace{N_r^2 N_t^2}_{(66)} + \underbrace{2N_r^3 N_t^3}_{(67)} + M^{N_t}\left(\underbrace{2N_t^2 N_r + 4N_r N_t}_{(72)}\right) \quad (80)$$

$$\kappa_{\text{vb}}^{(A)} = \underbrace{N_t^2 N_r^2 + N_r(N_r N_t - 1)(2N_r N_t + N_r)}_{(62)}$$

$$+ \underbrace{N_r^2(1 + N_t(N_r - 1)) + N_r^3}_{(62)}$$

$$+ \underbrace{4N_r(N_t N_r - 1) + 2N_r + N_r N_t}_{(63)}$$

$$+ \underbrace{N_r^2 N_t^2(N_r(N_t - 1) - 1)}_{(64)}$$

$$+ \underbrace{N_r^3 N_t^3 + N_r^2 N_t^2 + 2N_r^2 N_t^2(N_r N_t - 1)}_{(65)}$$

$$+ \underbrace{2N_r N_t + N_r N_t(N_r N_t - 1)}_{(66)}$$

$$+ \underbrace{2N_r^2 N_t^2 + 2N_r^2 N_t^2(N_r N_t - 1)}_{(67)}$$

$$+ M^{N_t}\left(\underbrace{4N_r N_t + 2N_r N_t^2}_{(72)}\right) \quad (81)$$

VIII. SIMULATION RESULTS

In this section we study the performance of the receiver algorithms proposed in the previous sections, namely the SPA-MAP (39), Gauss-MAP (51) and VB-MAP (72), and those from
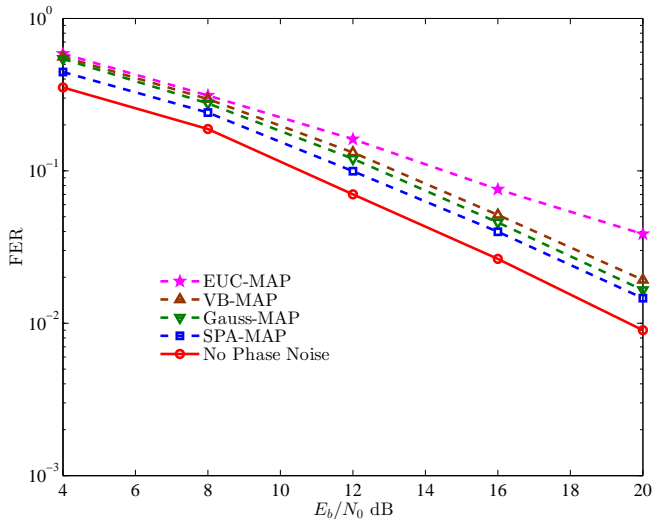
Fig. 4: FER performance of a 2×1 system using different receiver algorithms for code rate $= 3/4$ and $\sigma_{\mathrm{t}} = \sigma_{\mathrm{r}} = 4°$.

Fig. 5: SER performance of a 2×1 system for 16-QAM transmission using different receiver algorithms for $\sigma_{\mathrm{t}} = \sigma_{\mathrm{r}} = 4°$.

prior work, namely the EUC-MAP (55). The error performance curve corresponding to the case of no phase noise is used as the benchmark.

For implementing the Gauss-MAP, VB-MAP and EUC-MAP, the EKS is used to compute the estimate of the phase noise in each link and its respective variance. These estimates and their variances are used by the detector to compute the a posteriori symbol pmf, which we denote as $P_{\mathbf{c}}(\mathbf{c}_k)$. For the Gauss-MAP, EUC-MAP and VB-MAP, $P_{\mathbf{c}}(\mathbf{c}_k)$ is computed using (51), (55) and (72), respectively. In uncoded transmission, the detector computes the symbol average and its variance as $\mathbb{E}_{P_{\mathbf{c}}}\{\mathbf{c}_k\} = \underline{\mathbf{c}}_k$, $\mathrm{Var}_{P_{\mathbf{c}}}\{\mathbf{c}_k\} = \sigma_{\mathbf{c}}^2$. These symbol statistics are conveyed back to the EKS, and we perform 2 iterations between the smoother and the detector. The transmitted symbols are decided as $\hat{\mathbf{c}}_k = \arg\max_{\mathbf{c}_k} P_{\mathbf{c}}(\mathbf{c}_k)$. In coded transmission, the symbol pmf computed by the detector is used by the decoder for computing the bit LLRs, and in all, $L_{\mathrm{Det}} = 2$ global iterations are performed between the detector and the decoder. After the maximum number of iterations is reached, $P_{\mathbf{c}}(\mathbf{c}_k)$ is used by the decoder to make hard decisions on the information bits. The number of decoder iterations per global iteration is set to $L_{\mathrm{Dec}} = 30$. In coded transmissions, the EUC-SPA corresponds to the turbo-synchronization algorithm for MIMO systems proposed in [15], [35].

We consider data transmission in a strong phase noise scenario that corresponds to $\sigma_{\mathrm{t}} = \sigma_{\mathrm{r}} = 4°$ [8]. The channel is considered to be Rayleigh fading, and is assumed to be known (estimated) and constant over a data frame. Different channel realizations are generated for each data frame. In uncoded data transmission, the length of a data frame is $L = 10000$ symbols, and we consider the transmission of binary phase-shift keying (BPSK) symbols unless otherwise stated. In coded data transmission, we consider that the information bits are encoded with rate-$1/2$ and rate-$3/4$ LDPC code of length $L = 4000$, and the coded symbols are distributed over all the transmit antennas – i.e., the coded symbols at odd and even positions in the codeword (or the frame) are demultiplexed into different transmit antennas. For both transmission modes, we place 10 consecutive pilot symbols at the beginning of each frame, and 1 pilot symbol every 20 data symbols yielding a pilot density of around 5.1%.
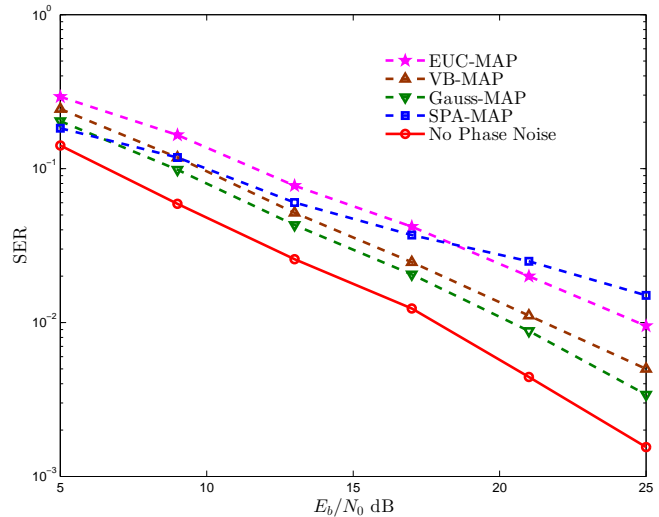
First, we investigate the performances of the proposed algorithms for different values of N_{t} and N_{r}. In Fig. 2, the bit error probability (BER) performance of the proposed algorithms is illustrated for 2×2 and 4×4 systems for uncoded transmission. For the 2×2 system, we note that all the proposed algorithms outperform the EUC-MAP for all values of SNR per bit (E_{b}/N_0). We observe that the SPA-MAP performs better than the Gauss-MAP and the VB-MAP especially for low values of E_{b}/N_0 by around 1 dB. For high E_{b}/N_0 values, it can be seen that the Gauss-MAP and SP-MAP render almost similar performance.

As both N_{t} and N_{r} are increased to 4, the gap between the proposed algorithms and the EUC-MAP widens. Furthermore, we observe that the gap in the performance between the benchmark algorithm, the proposed algorithms and the EUC-MAP also increases. In particular, the gap between the proposed algorithms and the EUC-SPA is around 2 dB for high values of E_{b}/N_0. This can be attributed to higher amplitude distortions due to phase noise experienced by the transmitted symbols as the number of antennas increases [16]. The superior performance of SPA-MAP confirms that the approximations made in Sec. IV are meaningful.

Next, in Fig. 3, we investigate the frame error rate (FER) performance of the proposed algorithms for the coded transmission mode considering rate-$1/2$ LDPC code and a 2×1 system. We observe that the SPA-MAP performs better than all the other algorithms. Specifically, we observe that the SPA-MAP outperforms the second-best algorithm (Gauss-MAP) by 0.8 dB.

In Fig. 4, we evaluate the FER performance of all algorithms for rate-$3/4$ LDPC code considering a 2×1 system in order to study the dependence of the performance of the algorithms on the code rate. We observe that the SPA-MAP outperforms the second-best algorithm (Gauss-MAP) by around 0.4 dB, and the gap in the performance between SPA-MAP and the EKS-based detectors decreases with increasing code rate (as compared to Fig. 3). This is because the pilot symbols have higher energy when the code rate is higher and are thus more reliable resulting in a better performance of the EKS.

We now analyze the performance of the algorithms for different constellation sizes. The symbol error rate (SER) performance
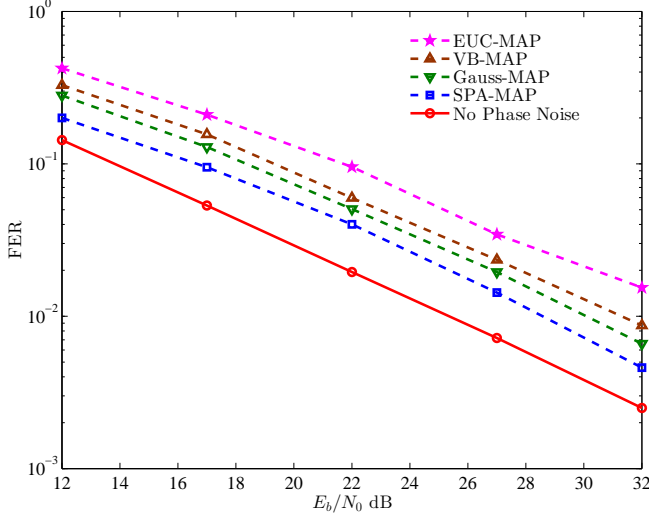
Fig. 6: FER performance of a 2×1 system for 16-QAM transmission, for code rate $= 3/4$ using different receiver algorithms for $\sigma_t = \sigma_r = 4°$.

TABLE I: Computational Complexity of the Receivers for $N_t = 2, N_r = 1, M = 2, r_c = .5, w_c = 3, w_r = 6$.

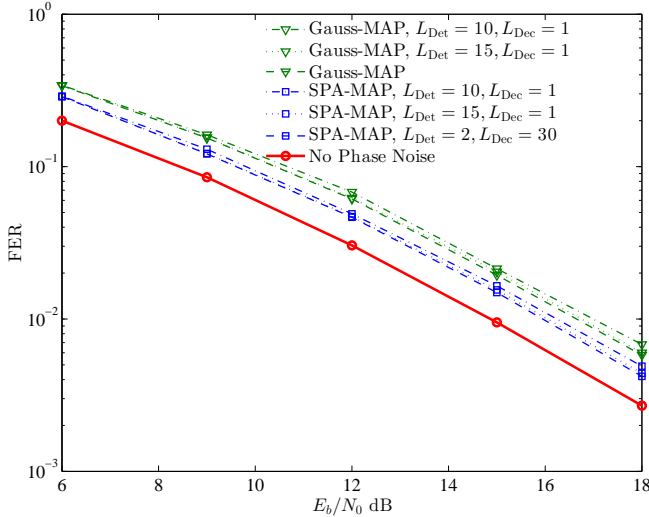| $L_{\det}, L_{\mathrm{dec}}$	SPA-MAP	Gauss-MAP	VB-MAP	EUC-MAP
10,1	3314	3491	3406	3382
15,1	4825	4980	4923	4902
2,30	3376	3538	3475	3424

Fig. 7: FER performance of a 2×1 system using SPA-MAP and Gauss-MAP for $r_c = 1/2$, $\sigma_t = \sigma_r = 4°$ for different $L_{\det}, L_{\mathrm{dec}}$ values.

of the algorithms is presented in Fig. 5 for uncoded 16-QAM transmission. We observe that the Gauss-MAP and the VB-MAP outperform the SPA-MAP significantly. This owes to the approximation of the Tikhonov mixture to a single mode Tikhonov pdf in (24) using uniform a priori information for the transmitted symbols in $P_d^{(c)}(\mathbf{c}_k)$. As stated before, this approximation can be highly erroneous for large non-equal energy constellations particularly when reliable a priori information of the transmitted symbols is not available to the detector. However, for the case of coded transmission considering rate-$3/4$ LDPC code in Fig. 6, the FER performance of the SPA-MAP algorithm is seen to be superior to all the other algorithms. This is because the LDPC decoder provides a more reliable a priori information in $P_d^{(c)}(\mathbf{c}_k)$ of the transmitted symbols rendering the single mode Gaussian pdf approximation more accurate.

Finally, in order analyze the complexity of the receivers, we study the case where we iterate between the detector and the decoder by setting $L_{\mathrm{Dec}} = 1$ for rate-$1/2$ LDPC code of length $L = 4000$, $\sigma_t = \sigma_r = 4°$ and a pilot density of around 5.1%. For the SNRs considered, we observe that $L_{\det} \approx 15$ iterations are required for convergence for the EKF-based detectors and SPA-MAP. We compare the computational complexity of the receivers in Table I for different values of $L_{\mathrm{Dec}}, L_{\mathrm{Det}}$ such that similar FER performances are attained. For the purpose of clarity, in Fig. 7 we only demonstrate the performance of the SPA-MAP and Gauss-MAP for different values of L_{dec} and $L_{\det}$. We can see that when $L_{\mathrm{dec}} = 1, L_{\det} = 15$, the performance of the receiver is the same as that when $L_{\mathrm{dec}} = 30, L_{\det} = 2$, but the complexity is higher in the former case. However, with a small loss in performance, choosing $L_{\mathrm{dec}} = 1, L_{\det} = 10$ gives an advantage in terms of complexity with respect to the case where $L_{\mathrm{dec}} = 30, L_{\det} = 2$. Clearly, the optimal choice of $L_{\mathrm{dec}}, L_{\det}$ (in terms of complexity) depends on $N_t, N_r, M, r_c, w_c, w_r$.

IX. CONCLUSIONS

In this paper, we derived the optimum MAP symbol detector that involves the joint estimation of the a posteriori phase noise pdf and data detection. The optimum receiver structure is seen to be intractable and unimplementable in practice. In this regard, we proposed three suboptimal, low-complexity algorithms that were observed to outperform the other algorithms available in the literature, which are considered in this work. In particular, the receiver based on the sum-product algorithm, SPA-MAP, was found to perform better than all the other algorithms for both uncoded and coded transmission of BPSK symbols. For higher-order constellations (16-QAM), the algorithms based on the smoother-detector structure, Gauss-MAP, and the variational Bayesian framework, VB-MAP, were observed to perform the best in the case of uncoded transmission. However, for coded 16-QAM transmission, the SPA-MAP algorithm was seen to be superior to all the other algorithms considered.

APPENDIX A
DERIVATION OF THE SPA MESSAGES AND COMPUTATION OF THEIR PARAMETERS

The message $p_f^{(\theta)}(\boldsymbol{\Theta}_k)$ is derived for the case of the Wiener phase noise process by evaluating (32) using the approximation [8, eq. (42)]

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_0^{2\pi} e^{\Re[ze^{-j\varphi}]} e^{\frac{-(\varphi-\phi)^2}{2\sigma^2}} \mathrm{d}\varphi \; \tilde{\propto} \; \exp\left\{ \Re\left[\frac{ze^{-j\phi}}{1+|z|\sigma^2} \right] \right\}$$
(82)

for $z \in \mathbb{C}$, $\sigma^2 \in \mathbb{R}^+$, and $\varphi, \phi \in \mathbb{R}$. We first evaluate

$$A_{\mathrm{temp}}^{(1)}$$
$$= \int_0^{2\pi} \exp\left\{ \Re\left[(y_{k-1}^{(1)} e^{-j\theta_{t,k-1}^{(1)}} + y_{k-1}^{(2)} e^{-j\theta_{t,k-1}^{(2)}}) e^{-j\theta_{r,k-1}^{(1)}} \right. \right.$$
$$\left. \left. - y_{k-1}^{(3)} e^{-j(\theta_{t,k-1}^{(1)} - \theta_{t,k-1}^{(2)})} \right] \right\}$$
$$\cdot p_\Delta(\theta_{r,k}^{(1)} - \theta_{r,k-1}^{(1)}) \mathrm{d}\theta_{r,k-1}^{(1)}$$
(83)
$$= \exp\left\{ -\Re\left[y_{k-1}^{(3)} e^{-j(\theta_{t,k-1}^{(1)} - \theta_{t,k-1}^{(2)})} \right] \right\}$$
$$\cdot \int_0^{2\pi} \exp\left\{ \Re\left[\left(y_{k-1}^{(1)} e^{-j\theta_{t,k-1}^{(1)}} + y_{k-1}^{(2)} e^{-j\theta_{t,k-1}^{(2)}} \right) e^{-j\theta_{r,k-1}^{(1)}} \right] \right\}$$

$$\cdot\, p_\Delta(\theta_{\mathrm{r},k}^{(1)} - \theta_{\mathrm{r},k-1}^{(1)})\mathrm{d}\theta_{\mathrm{r},k-1}^{(1)}$$

$$\propto \exp\left\{-\Re\left[y_{k-1}^{(3)}e^{-\jmath(\theta_{\mathrm{t},k-1}^{(1)}-\theta_{\mathrm{t},k-1}^{(2)})}\right]\right\}$$

$$\cdot \exp\left\{\Re\left[\frac{\left(y_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(1)}} + y_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right)e^{-\jmath\theta_{\mathrm{r},k}^{(1)}}}{1+\sigma_{\mathrm{r}}^2\left|y_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(1)}} + y_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right|}\right]\right\} \quad (84)$$

$$\approx \exp\left\{-\Re\left[y_{k-1}^{(3)}e^{-\jmath(\theta_{\mathrm{t},k-1}^{(1)}-\theta_{\mathrm{t},k-1}^{(2)})}\right]\right\}$$

$$\cdot \exp\left\{\Re\left[\frac{\left(y_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(1)}} + y_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right)}{1+\sigma_{\mathrm{r}}^2\left(|y_{k-1}^{(1)}| + |y_{k-1}^{(2)}|\right)}e^{-\jmath\theta_{\mathrm{r},k}^{(1)}}\right]\right\}. \quad (85)$$

In (84), we used the approximation in (82), and to obtain the result in (85) we applied the approximation

$$\left|y_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(1)}} + y_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right|$$

$$= \left|\left|y_{k-1}^{(1)}\right|e^{\jmath(\angle y_{k-1}^{(1)}-\theta_{\mathrm{t},k-1}^{(1)}-\theta_{\mathrm{r},k-1}^{(1)})}\right.$$

$$\left. + \left|y_{k-1}^{(2)}\right|e^{\jmath(\angle y_{k-1}^{(2)}-\theta_{\mathrm{r},k-1}^{(1)}-\theta_{\mathrm{t},k-1}^{(2)})}\right| \quad (86)$$

$$\approx \left|y_{k-1}^{(1)}\right| + \left|y_{k-1}^{(2)}\right|, \quad (87)$$

where it is assumed that $\angle y_{k-1}^{(2)} - \theta_{\mathrm{r},k-1}^{(1)} - \theta_{\mathrm{t},k-1}^{(2)}$ and $\angle y_{k-1}^{(1)} - \theta_{\mathrm{r},k-1}^{(1)} - \theta_{\mathrm{t},k-1}^{(1)}$ are very small, i.e., we assume that the difference between the phase noise in each link and its (predicted and updated) estimate is small. Now, define

$$\tilde{y}_{k-1}^{(1)} \triangleq \frac{y_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{r},k}^{(1)}}}{1+\sigma_{\mathrm{r}}^2\left(|y_{k-1}^{(1)}| + |y_{k-1}^{(2)}|\right)}$$

$$\tilde{y}_{k-1}^{(2)} \triangleq \frac{y_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{r},k}^{(1)}}}{1+\sigma_{\mathrm{r}}^2\left(|y_{k-1}^{(1)}| + |y_{k-1}^{(2)}|\right)}. \quad (88)$$

Then compute

$$A_{\mathrm{temp}}^{(2)} \triangleq \int_0^{2\pi} A_{\mathrm{temp}}^{(1)} p_\Delta(\theta_{\mathrm{t},k}^{(1)} - \theta_{\mathrm{t},k-1}^{(1)})\mathrm{d}\theta_{\mathrm{t},k-1}^{(1)}$$

$$= \int_0^{2\pi} \exp\left\{\Re\left[\tilde{y}_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(1)}} + \tilde{y}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right.\right.$$

$$\left.\left. - y_{k-1}^{(3)}e^{-\jmath(\theta_{\mathrm{t},k-1}^{(1)}-\theta_{\mathrm{t},k-1}^{(2)})}\right]\right\}$$

$$\cdot\, p_\Delta(\theta_{\mathrm{t},k}^{(1)} - \theta_{\mathrm{t},k-1}^{(1)})\mathrm{d}\theta_{\mathrm{t},k-1}^{(1)}$$

$$= \exp\left\{\Re\left[\tilde{y}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right]\right\}$$

$$\cdot \int \exp\left\{\Re\left[\left(\tilde{y}_{k-1}^{(1)} - y_{k-1}^{(3)}e^{\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right)e^{-\jmath\theta_{\mathrm{t},k-1}^{(1)}}\right]\right\}$$

$$\cdot\, p_\Delta(\theta_{\mathrm{t},k}^{(1)} - \theta_{\mathrm{t},k-1}^{(1)})\mathrm{d}\theta_{\mathrm{t},k-1}^{(1)}$$

$$\propto \exp\left\{\Re\left[\tilde{y}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right]\right\}$$

$$\cdot \exp\left\{\Re\left[\frac{\left(\tilde{y}_{k-1}^{(1)} - y_{k-1}^{(3)}e^{\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right)e^{-\jmath\theta_{\mathrm{t},k}^{(1)}}}{1+\sigma_{\mathrm{t}}^2\left|\tilde{y}_{k-1}^{(1)} - y_{k-1}^{(3)}e^{\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right|}\right]\right\} \quad (89)$$

$$\approx \exp\left\{\Re\left[\tilde{y}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right]\right\}$$

$$\cdot \exp\left\{\Re\left[\frac{\left(\tilde{y}_{k-1}^{(1)} - y_{k-1}^{(3)}e^{\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right)e^{-\jmath\theta_{\mathrm{t},k}^{(1)}}}{1+\sigma_{\mathrm{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right| - \left|y_{k-1}^{(3)}\right|\right|}\right]\right\}, \quad (90)$$

where in (89), the approximation from (82) is used. In (90), we apply

$$\left|\tilde{y}_{k-1}^{(1)} - y_{k-1}^{(3)}e^{\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right|$$

$$= \left|\left|\tilde{y}_{k-1}^{(1)}\right|e^{\jmath\angle\tilde{y}_{k-1}^{(1)}} - \left|y_{k-1}^{(3)}\right|e^{\jmath\angle y_{k-1}^{(3)}}e^{\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right|$$

$$\simeq \left|\left|\tilde{y}_{k-1}^{(1)}\right|e^{\jmath\angle\tilde{y}_{k-1}^{(1)}} - \left|y_{k-1}^{(3)}\right|e^{\jmath(\angle y_{k-1}^{(1)}-\angle y_{k-1}^{(2)}+\theta_{\mathrm{t},k-1}^{(2)})}\right|$$

$$= \left|\left|\tilde{y}_{k-1}^{(1)}\right|e^{\jmath\angle\tilde{y}_{k-1}^{(1)}} - \left|y_{k-1}^{(3)}\right|e^{\jmath(\angle\tilde{y}_{k-1}^{(1)}-\angle y_{k-1}^{(2)}+\theta_{\mathrm{r},k}^{(1)}+\theta_{\mathrm{t},k-1}^{(2)})}\right|$$

$$\approx \left|\left|\tilde{y}_{k-1}^{(1)}\right|e^{\jmath\angle\tilde{y}_{k-1}^{(1)}} - \left|y_{k-1}^{(3)}\right|e^{\jmath\angle\tilde{y}_{k-1}^{(1)}}\right| = \left|\left|\tilde{y}_{k-1}^{(1)}\right| - \left|y_{k-1}^{(3)}\right|\right|, \quad (91)$$

where in (91), it is considered that $\theta_{\mathrm{r},k}^{(1)} + \theta_{\mathrm{t},k-1}^{(2)} - \angle y_{k-1}^{(2)}$ is very small[2]. This is reasonable, since practically strong phase noise refers to scenarios where $\sigma_\Delta^2 \approx 10^{-3} - 10^{-2}$ rad^2 or $\sigma_\Delta \approx 2 - 6°$ and weak-to-moderate scenarios correspond to $\sigma_\Delta^2 \approx 10^{-5} - 10^{-4}$ rad^2 or $\sigma_\Delta \approx 0.18 - 0.6°$.

Finally, we compute

$$A_{\mathrm{temp}}^{(3)} \triangleq \int_0^{2\pi} A_{\mathrm{temp}}^{(2)} p_\Delta(\theta_{\mathrm{t},k}^{(2)} - \theta_{\mathrm{t},k-1}^{(2)})\mathrm{d}\theta_{\mathrm{t},k-1}^{(2)}$$

$$= \int_0^{2\pi} \exp\left\{\Re\left[\tilde{y}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right]\right\}$$

$$\cdot \exp\left\{\Re\left[\frac{\left(\tilde{y}_{k-1}^{(1)} - y_{k-1}^{(3)}e^{\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right)}{1+\sigma_{\mathrm{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right| - \left|y_{k-1}^{(3)}\right|\right|}e^{-\jmath\theta_{\mathrm{t},k}^{(1)}}\right]\right\}$$

$$\cdot\, p_\Delta(\theta_{\mathrm{t},k}^{(2)} - \theta_{\mathrm{t},k-1}^{(2)})\mathrm{d}\theta_{\mathrm{t},k-1}^{(2)}$$

$$= \exp\left\{\Re\left[\frac{\tilde{y}_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k}^{(1)}}}{1+\sigma_{\mathrm{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right| - \left|y_{k-1}^{(3)}\right|\right|}\right]\right\}$$

$$\cdot \int_0^{2\pi} \exp\left\{\Re\left[\tilde{\tilde{y}}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k-1}^{(2)}}\right]\right\} p_\Delta(\theta_{\mathrm{t},k}^{(2)} - \theta_{\mathrm{t},k-1}^{(2)})\mathrm{d}\theta_{\mathrm{t},k-1}^{(2)}$$

$$\propto \exp\left\{\Re\left[\frac{\tilde{y}_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k}^{(1)}}}{1+\sigma_{\mathrm{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right| - \left|y_{k-1}^{(3)}\right|\right|} + \frac{\tilde{\tilde{y}}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k}^{(2)}}}{1+\sigma_{\mathrm{t}}^2\left|\tilde{\tilde{y}}_{k-1}^{(2)}\right|}\right]\right\}$$

$$\approx \exp\left\{\Re\left[\frac{\tilde{y}_{k-1}^{(1)}e^{-\jmath\theta_{\mathrm{t},k}^{(1)}}}{1+\sigma_{\mathrm{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right| - \left|y_{k-1}^{(3)}\right|\right|}\right.\right.$$

$$\left.\left. + \frac{\tilde{\tilde{y}}_{k-1}^{(2)}e^{-\jmath\theta_{\mathrm{t},k}^{(2)}}}{1+\sigma_{\mathrm{t}}^2\left|\left|\tilde{y}_{k-1}^{(2)}\right| - \left|y_{k-1}^{(3)}\right|\right|}\right]\right\}, \text{ where} \quad (92)$$

$$\tilde{\tilde{y}}_{k-1}^{(2)} \triangleq \tilde{y}_{k-1}^{(2)} - \frac{y_{k-1}^{(3)\,*}e^{\jmath\theta_{\mathrm{t},k}^{(1)}}}{1+\sigma_{\mathrm{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right| - \left|y_{k-1}^{(3)}\right|\right|}.$$

To obtain (92), we apply the approximation from (82), and further we apply an approximation similar to that used in (91).

---

[2]Here, we assume that the difference between the phase noise in each link and its (predicted and updated) estimate is small in the sense that we expect the phase noise to be tracked by the algorithm with reasonable accuracy. This assumption works well in the phase noise context. Note that a similar assumption is used to perform linearization about a phase noise estimate in an EKS

Observe that the message $p_{\text{f}}^{(\theta)}(\boldsymbol{\Theta}_k) = A_{\text{temp}}^{(3)}$, and hence

$$
p_{\text{f}}^{(\theta)}(\boldsymbol{\Theta}_k)
$$

$$
\approx \exp\left\{\Re\left[\frac{\tilde{y}_{k-1}^{(1)}e^{-\jmath\theta_{\text{t},k}^{(1)}}}{1+\sigma_{\text{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right|-\left|y_{k-1}^{(3)}\right|\right|} + \frac{\tilde{y}_{k-1}^{(2)}e^{-\jmath\theta_{\text{t},k}^{(2)}}}{1+\sigma_{\text{t}}^2\left|\left|\tilde{y}_{k-1}^{(2)}\right|-\left|y_{k-1}^{(3)}\right|\right|}\right.\right.
$$

$$
\left.\left.-\frac{y_{k-1}^{(3)}e^{-\jmath(\theta_{\text{t},k}^{(1)}-\theta_{\text{t},k}^{(2)})}}{\left(1+\sigma_{\text{t}}^2\left|\left|\tilde{y}_{k-1}^{(2)}\right|-\left|y_{k-1}^{(3)}\right|\right|\right)\left(1+\sigma_{\text{t}}^2\left|\left|\tilde{y}_{k-1}^{(1)}\right|-\left|y_{k-1}^{(3)}\right|\right|\right)}\right]\right\}
$$

$$
\triangleq \exp\left\{\Re\left[(a_{\text{f},k}^{(1,1)}e^{-\jmath\theta_{\text{t},k}^{(1)}} + a_{\text{f},k}^{(2,1)}e^{-\jmath\theta_{\text{t},k}^{(2)}})e^{-\jmath\theta_{\text{r},k}^{(1)}} - \tilde{a}_{\text{f},k}^{(1,2)}\right.\right.
$$

$$
\left.\left.\cdot e^{-\jmath(\theta_{\text{t},k}^{(1)}-\theta_{\text{t},k}^{(2)})}\right]\right\}. \tag{93}
$$

From the result in (93), we arrive at the forward recursions presented in (34). Computation of the message $p_{\text{b}}^{(\theta)}(\boldsymbol{\Theta}_k)$ to determine the backward parameter update equations in (37) proceeds similarly.

For computing the message $P_{\text{u}}^{(c)}(\mathbf{c}_k)$ in (39), define

$$
A_{\text{temp}}^{(4)} \triangleq \int \exp\left\{\Re\left[(z_k^{(1,1)}e^{-\jmath\theta_{\text{t},k}^{(1)}} + z_k^{(2,1)}e^{-\jmath\theta_{\text{t},k}^{(2)}})e^{-\jmath\theta_{\text{r},k}^{(1)}}\right.\right.
$$

$$
\left.\left.- \tilde{z}_k^{(1,2)}e^{-\jmath(\theta_{\text{t},k}^{(1)}-\theta_{\text{t},k}^{(2)})}\right]\right\}d\boldsymbol{\Theta}_k. \tag{94}
$$

The integral in (94) is evaluated as

$$
A_{\text{temp}}^{(4)} = \int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\exp\left\{\Re\left[(z_k^{(1,1)}e^{-\jmath\theta_{\text{t},k}^{(1)}} + z_k^{(2,1)}e^{-\jmath\theta_{\text{t},k}^{(2)}})\right.\right.
$$

$$
\left.\left.\cdot e^{-\jmath\theta_{\text{r},k}^{(1)}} - \tilde{z}_k^{(1,2)}e^{-\jmath(\theta_{\text{t},k}^{(1)}-\theta_{\text{t},k}^{(2)})}\right]\right\}d\theta_{\text{r},k}^{(1)}d\theta_{\text{t},k}^{(2)}d\theta_{\text{t},k}^{(1)}
$$

$$
= \int_0^{2\pi}\int_0^{2\pi}\exp\left\{-\Re\left[\tilde{z}_k^{(1,2)}e^{-\jmath(\theta_{\text{t},k}^{(1)}-\theta_{\text{t},k}^{(2)})}\right]\right\}
$$

$$
\cdot\int_0^{2\pi}\exp\left\{\Re\left[\left(z_k^{(1,1)}e^{-\jmath\theta_{\text{t},k}^{(1)}} + z_k^{(2,1)}e^{-\jmath\theta_{\text{t},k}^{(2)}}\right)e^{-\jmath\theta_{\text{r},k}^{(1)}}\right]\right\}
$$

$$
d\theta_{\text{r},k}^{(1)}d\theta_{\text{t},k}^{(2)}d\theta_{\text{t},k}^{(1)}
$$

$$
\propto \int_0^{2\pi}\int_0^{2\pi}\exp\left\{-\Re\left[\tilde{z}_k^{(1,2)}e^{-\jmath(\theta_{\text{t},k}^{(1)}-\theta_{\text{t},k}^{(2)})}\right]\right\}
$$

$$
\cdot\mathrm{I}_0\left(\left|z_k^{(1,1)}e^{-\jmath\theta_{\text{t},k}^{(1)}} + z_k^{(2,1)}e^{-\jmath\theta_{\text{t},k}^{(2)}}\right|\right)d\theta_{\text{t},k}^{(2)}d\theta_{\text{t},k}^{(1)} \tag{95}
$$

$$
\approx \int_0^{2\pi}\int_0^{2\pi}\exp\left\{-\Re\left[\tilde{z}_k^{(1,2)}e^{-\jmath(\theta_{\text{t},k}^{(1)}-\theta_{\text{t},k}^{(2)})}\right]\right\}
$$

$$
\cdot\mathrm{I}_0\left(\left|\left|z_k^{(1,1)}\right| + \left|z_k^{(2,1)}\right|\right|\right)d\theta_{\text{t},k}^{(2)}d\theta_{\text{t},k}^{(1)} \tag{96}
$$

$$
\propto \mathrm{I}_0\left(\left|\tilde{z}_k^{(1,2)}\right|\right)\mathrm{I}_0\left(\left|\left|z_k^{(1,1)}\right| + \left|z_k^{(2,1)}\right|\right|\right), \tag{97}
$$

where in (95) the result follows from (82), and the result in (96) is obtained by applying the approximation in (86). Further, applying the Tikhonov normalization constant, the result in (97) is obtained. The result in (97) is used in (94) to yield $P_{\text{u}}^{(c)}(\mathbf{c}_k)$ in (39).

The generalizations presented in (41) and (43) are obtained by first identifying that $p_{\text{d}}^{(\theta)}(\boldsymbol{\Theta}_k)$ fully factorizes in terms of the receive antenna index. Further, the Gaussian pdf that minimizes the KL divergence with respect to $p_{\text{d}}^{(\theta)}(\boldsymbol{\Theta}_k)$ is obtained by performing moment matching with each of its factors [22]. Following this, the remaining steps proceed similarly as presented in (83)-(97).

APPENDIX B
DERIVATION OF THE APPROXIMATE MAP DETECTOR BASED ON THE SMOOTHER–DETECTOR STRUCTURE

The approximate MAP detector based on the smoother-detector structure in (50) is rewritten as

$$
\hat{\mathbf{c}}_k \propto \arg\max_{\mathbf{c}_k}\exp\left\{-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0}\right\}
$$

$$
\int_{\boldsymbol{\Theta}_k}\exp\left\{\Re\left[\frac{2}{N_0}r_k^{(1)}c_{\text{h},k}^{(1,1)*}e^{-\jmath\theta_k^{(1,1)}} + \frac{2}{N_0}r_k^{(2)}c_{\text{h},k}^{(2,1)*}\right.\right.
$$

$$
\left.\left.\cdot e^{-\jmath\theta_k^{(2,1)}} - \frac{2}{N_0}c_{\text{h},k}^{(2,1)}c_{\text{h},k}^{(1,1)*}e^{-\jmath(\theta_k^{(1,1)}-\theta_k^{(2,1)})}\right]\right\}
$$

$$
\mathcal{N}(\boldsymbol{\Theta}_k; \widehat{\boldsymbol{\Theta}}_k, \hat{\mathbf{P}}_k)d\boldsymbol{\Theta}_k \tag{98}
$$

$$
\approx \arg\max_{\mathbf{c}_k}\exp\left\{-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0}\right\}
$$

$$
\int_{\boldsymbol{\Theta}_k}\exp\left\{\Re\left[\frac{2}{N_0}r_k^{(1)}c_{\text{h},k}^{(1,1)*}e^{-\jmath\theta_k^{(1,1)}} + \frac{2}{N_0}r_k^{(2)}c_{\text{h},k}^{(2,1)*}\right.\right.
$$

$$
\left.\left.\cdot e^{-\jmath\theta_k^{(2,1)}} - \frac{2}{N_0}c_{\text{h},k}^{(2,1)}c_{\text{h},k}^{(1,1)*}e^{-\jmath(\theta_k^{(1,1)}-\theta_k^{(2,1)})}\right]\right\}
$$

$$
\cdot\exp\left\{\Re\left[\frac{e^{\jmath\hat{\theta}_k^{(1,1)}}}{P_{k,1}^{(1,1)}}e^{-\jmath\theta_k^{(1,1)}} + \frac{e^{\jmath\hat{\theta}_k^{(2,1)}}}{P_{k,1}^{(2,1)}}e^{-\jmath\theta_k^{(2,1)}}\right.\right.
$$

$$
\left.\left.- \left|\tilde{u}_k^{(1,2)}\right|e^{-\jmath(\theta_k^{(1,1)}-\theta_k^{(2,1)})}\right]\right\}d\boldsymbol{\Theta}_k \tag{99}
$$

$$
\approx \arg\max_{\mathbf{c}_k}B_{\text{temp}}^{(1)}, \tag{100}
$$

where, in (99), we approximate the bivariate Gaussian pdf as a bivariate Tikhonov distribution [25]. This is a reasonable approximation for variances of around 10^{-1} or smaller associated with $\theta_k^{(1,1)}, \theta_k^{(2,1)}$ [25], which encompass practical scenarios of weak-to-strong phase noise. Note that a similar approximation holds in the case of univariate Gaussian and Tikhonov pdfs and has been widely used in the literature [10].

We simplify $B_{\text{temp}}^{(1)}$ as

$$
B_{\text{temp}}^{(1)}
$$

$$
\triangleq \exp\left\{-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0}\right\}\exp\left\{\Re\left[u_k^{(2,1)}e^{-\jmath\theta_k^{(2,1)}}\right]\right\}
$$

$$
\cdot\int_{\boldsymbol{\Theta}_k}\exp\left\{\Re\left[u_k^{(1,1)}e^{-\jmath\theta_k^{(1,1)}} - \tilde{u}_k^{(1,2)}e^{-\jmath(\theta_k^{(1,1)}-\theta_k^{(2,1)})}\right]\right\}d\boldsymbol{\Theta}_k \tag{101}
$$

$$
\propto \exp\left\{-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0}\right\}\mathrm{I}_0\left(|u_k^{(1,1)} - \tilde{u}_k^{(1,2)}e^{-\jmath\theta_k^{(2,1)}}|\right)
$$

$$
\cdot\int_{\boldsymbol{\Theta}_k}\exp\left\{\Re\left[u_k^{(2,1)}e^{-\jmath\theta_k^{(2,1)}}\right]\right\}d\boldsymbol{\Theta}_k \tag{102}
$$

$$
\approx \exp\left\{-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0}\right\}\mathrm{I}_0\left(||u_k^{(1,1)}| - |\tilde{u}_k^{(1,2)}||\right)
$$

$$
\cdot\mathrm{I}_0\left(|u_k^{(2,1)}|\right) \tag{103}
$$

$$
\approx \exp\left\{-\frac{|c_{\text{h},k}^{(1,1)}|^2 + |c_{\text{h},k}^{(2,1)}|^2}{N_0}\right\}\mathrm{I}_0\left(|u_k^{(1,1)}| + |u_k^{(2,1)}| - |\tilde{u}_k^{(1,2)}|\right), \tag{104}
$$

where we use the definitions from (52) in (101) and the Tikhonov pdf normalization constant in (102). In (103), we use the approximation from (91) where it is assumed that the difference between the predicted and updated estimates and the states is very small. Finally, in (104) we approximate $I_0(x) \approx e^x$ for large x and assume $|u_k^{(1,1)}| - |\tilde{u}_k^{(1,2)}| > 0$. The generalization presented in (54) is a straightforward extension of the computations from (98)-(104) for arbitrary values of N_t and N_r.

APPENDIX C
DERIVATION OF THE FACTORIZED PDFS FOR THE VB FRAMEWORK

From (70), the factorized pdf of $\bar{\Theta}$ is derived as

$$q_{\boldsymbol{\theta}}(\bar{\Theta}) = C_{\boldsymbol{\theta}} p(\bar{\Theta}) \prod_{k=1}^{L} \exp\left\{ C_{\text{temp}}^{(2)} \right\}, \text{ where }$$

$$C_{\text{temp}}^{(2)} \triangleq \mathbb{E}_{q_{\mathbf{c}}} \left\{ \log p(\mathbf{r}_k | \mathbf{c}_k, \Theta_k) \right\}$$

$$= \mathbb{E}_{q_{\mathbf{c}}} \left\{ \log \prod_{n=1}^{N_r} p(r_k^{(n)} | \mathbf{c}_k, \Theta_k) \right\},$$

$$= -\frac{1}{N_0} \sum_{n=1}^{N_r} \mathbb{E}_{q_{\mathbf{c}}} \left(r_k^{(n)} r_k^{(n)*} \right.$$

$$- \sum_{m=1}^{N_t} \left(\left| c_{\text{h},k}^{(m,n)} \right|^2 + \sum_{\substack{l=1 \\ l \neq m}}^{N_t} c_{\text{h},k}^{(m,n)} c_{\text{h},k}^{(l,n)*} e^{\jmath \left(\theta_k^{(m,n)} - \theta_k^{(l,n)} \right)} \right)$$

$$\left. - r_k^{(n)*} \sum_{m=1}^{N_t} c_{\text{h},k}^{(m,n)} e^{\jmath \theta_k^{(m,n)}} - r_k^{(n)} \sum_{m=1}^{N_t} c_{\text{h},k}^{(m,n)*} e^{-\jmath \theta_k^{(m,n)}} \right)$$

$$= -\frac{1}{N_0} \sum_{n=1}^{N_r} \left(r_k^{(n)} r_k^{(n)*} + \sum_{m=1}^{N_t} \left(\left| \underline{c}_{\text{h},k}^{(m,n)} \right|^2 \right. \right.$$

$$+ \sum_{\substack{l=1 \\ l \neq m}}^{N_t} \underline{c}_{\text{h},k}^{(l,n)} \underline{c}_{\text{h},k}^{(m,n)*} e^{\jmath \left(\theta_k^{(m,n)} - \theta_k^{(m,n)} \right)} \right) - r_k^{(n)*} \sum_{m=1}^{N_t} \underline{c}_{\text{h},k}^{(m,n)}$$

$$\left. \cdot e^{\jmath \theta_k^{(m,n)}} - r_k^{(n)} \sum_{m=1}^{N_t} \underline{c}_{\text{h},k}^{(m,n)*} e^{-\jmath \theta_k^{(m,n)}} - \sum_{m=1}^{N_t} \text{Var}_{q_{\mathbf{c}}} c_{\text{h},k}^{(m,n)} \right). \tag{105}$$

$$= C_{\boldsymbol{\theta}} p(\bar{\Theta}) \prod_{k=1}^{L} \exp\left\{ \sum_{n=1}^{N_r} \frac{- \left| r_k^{(n)} - \sum_{m=1}^{N_t} \underline{c}_{\text{h},k}^{(m,n)} e^{\jmath \theta_k^{(m,n)}} \right|^2}{N_0} \right\}$$

$$= C_{\boldsymbol{\theta}} P(\bar{\Theta}) \prod_{k=1}^{L} p(\mathbf{r}_k | \Theta_k, \underline{\mathbf{c}}_k), \tag{106}$$

$$= C_{\boldsymbol{\theta}} p(\bar{\mathbf{r}}, \bar{\Theta} | \underline{\mathbf{c}}). \tag{107}$$

In (105), $\underline{c}_{\text{h},k}^{(m,n)}$ is defined as in (71), and for obtaining the result in (106) we apply the approximation that $\text{Var}_{q_{\mathbf{c}}} = \mathbb{E}_{q_{\mathbf{c}}} \left\{ c_{\text{h},k}^{(m,n)} c_{\text{h},k}^{(m,n)*} - \underline{c}_{\text{h},k}^{(m,n)} \underline{c}_{\text{h},k}^{(m,n)*} \right\}$ is very small. This implies that the estimation algorithm assumes that the a posteriori pmf of the transmitted symbols has a prominent peak. The quality of this approximation depends on the reliability of the a priori information available about the transmitted symbols. In particular, this approximation can be conjectured to be fairly accurate in coded systems. The constant $C_{\boldsymbol{\theta}}$ is the pdf normalizing factor such that $q_{\boldsymbol{\theta}}$ integrates to unity. Its value can be determined as

$$C_{\boldsymbol{\theta}} = \frac{1}{p(\bar{\mathbf{r}} | \underline{\mathbf{c}})}.$$

Plugging $C_{\boldsymbol{\theta}}$ in (107), the factorized pdf of $\bar{\Theta}$ is obtained as

$$q_{\boldsymbol{\theta}} = p(\bar{\Theta} | \bar{\mathbf{r}}, \underline{\mathbf{c}}). \tag{108}$$

From (70), the factorized pmf of $\bar{\mathbf{c}}$ is derived as

$$q_{\mathbf{c}} = C_{\mathbf{c}} \prod_{k=1}^{L} P(\mathbf{c}_k) \exp\left\{ C_{\text{temp}}^{(3)} \right\}, \text{ where } \tag{109}$$

$$C_{\text{temp}}^{(3)} \triangleq \int_{\boldsymbol{\theta}} q_{\boldsymbol{\theta}} \log P(\mathbf{r}_k | \mathbf{c}_k, \Theta_k) \mathrm{d}\Theta_k$$

$$= \mathbb{E}_{q_{\boldsymbol{\theta}}} \log \prod_{n=1}^{N_r} p(r_k^{(n)} | \mathbf{c}_k, \Theta_k)$$

$$= -\frac{1}{N_0} \sum_{n=1}^{N_r} \mathbb{E}_{q_{\boldsymbol{\theta}}} \left| r_k^{(n)} - \sum_{m=1}^{N_t} c_{\text{h},k}^{(m,n)} e^{\jmath \theta_k^{(k)}} \right|^2$$

$$= -\frac{1}{N_0} \sum_{n=1}^{N_r} \left\{ r_k^{(n)} r_k^{(n)*} \left(\left| c_{\text{h},k}^{(m,n)} \right|^2 \right. \right.$$

$$+ \sum_{m=1}^{N_t} \left(1 + \mathbb{E}_{q_{\boldsymbol{\theta}}} (\theta_k^{(m,n)} - \hat{\theta}_k^{(m,n)})^2 \right) + \sum_{\substack{l=1 \\ l \neq m}}^{N_t} c_{\text{h},k}^{(m,n)} c_{\text{h},k}^{(l,n)*}$$

$$\cdot e^{\jmath(\hat{\theta}_k^{(m,n)} - \hat{\theta}_k^{(l,n)})} (1 + \mathbb{E}_{q_{\boldsymbol{\theta}}} (\theta_k^{(m,n)} - \hat{\theta}_k^{(m,n)})(\theta_k^{(l,n)} - \hat{\theta}_k^{(l,n)}))$$

$$\left. - r_k^{(n)*} \sum_{m=1}^{N_t} c_{\text{h},k}^{(m,n)} (n) e^{\jmath \hat{\theta}_k^{(m,n)}} - r_k^{(n)} \sum_{m=1}^{N_t} c_{\text{h},k}^{(m,n)*} e^{-\jmath \hat{\theta}_k^{(m,n)}} \right\},$$

$$= -\frac{1}{N_0} \left\{ \sum_{n=1}^{N_r} \left| r_k^{(n)} - \sum_{m=1}^{N_t} c_{\text{h},k}^{(m,n)} e^{\jmath \hat{\theta}_k^{(m,n)}} \right|^2 - \sum_{m=1}^{N_t} \left| c_{\text{h},k}^{(m,n)} \right|^2 \right.$$

$$\left. \cdot P_{k,n}^{(m,m)} - \sum_{m=1}^{N_t} \sum_{\substack{l=1 \\ l \neq m}}^{N_t} c_{\text{h},k}^{(m,n)} c_{\text{h},k}^{(l,n)*} P_{k,n}^{(m,l)} e^{\jmath(\hat{\theta}_k^{(m,n)} - \jmath \hat{\theta}_k^{(l,n)})} \right\}. \tag{110}$$

Upon plugging the result from (110) into (109) we obtain the result in (72).

REFERENCES

[1] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge, U.K.: Cambridge Univ. Press, 2003.

[2] D. S. Baum, and H. Bölcskei, "Information-theoretic analysis of MIMO channel sounding," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7555-7577, Nov. 2011.

[3] N. Hadaschik, M. Dorpinghaus, A. Senst, O. Harmjanz, U. Kaufer, G. Ascheid, and H. Meyr, "Improving MIMO phase noise estimation by exploiting spatial correlations," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. 3, pp. 833-836, Mar. 2005.

[4] F. Bohagen, P. Orten, and G. E. Oien, "Design of capacity-optimal high-rank line-of-sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 790-804, Apr. 2007.

[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[6] R. Krishnan, M. R. Khanzadi, N. Krishnan, Y. Wu, A. Graell i Amat, T. Eriksson, and R. Schober, "Linear Massive MIMO Precoders in the Presence of Phase Noise – A Large-Scale Analysis," *IEEE Trans. Veh. Tech.*, accepted for publication. [Online]. Available: http://arxiv.org/abs/1501.05461

[7] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "A theoretical framework for soft information based synchronization in iterative (turbo) receivers," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, pp. 117-125, Apr. 2005.

[8] 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, Sept. 2005.

[9] M. Nissila, and S. Pasupathy, "Adaptive iterative detectors for phase-uncertain channels via variational bounding," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 716-725, Mar. 2009.

[10] R. Krishnan, M. R. Khanzadi, T. Eriksson, and T. Svensson, "Soft metrics and their performance analysis for optimal data detection in the presence of strong oscillator phase noise," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2385-2395, Jun. 2013.

[11] P. Worthen, and W. E Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 843-849, Feb. 2001.

[12] P. Y. Kam, S. S. Ng, and T. S. Ng, "Optimum symbol-by-symbol detection of uncoded digital data over the Gaussian channel with unknown carrier phase," *IEEE Trans. Commun.*, vol. 42, no. 8, pp. 2543-2552, Aug. 1994.

[13] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. 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.

[14] P. Almers, S. Wyne, F. Tufvesson, and A. F. Molisch, "Effect of random walk phase noise on MIMO measurements," *IEEE 61st VTC 2005-Spring*, pp. 141-145, 30 May-Jun. 2005.

[15] A. A. Nasir, H. Mehrpouyan, R. Schober, and Y. Hua, "Phase noise in MIMO Systems: Bayesian Cramer Rao bounds and soft-input estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2675-2692, May 2013.

[16] R. Krishnan, M. R. Khanzadi, L. Svensson, T. Eriksson, and T. Svensson, "Variational Bayesian framework for receiver design in the presence of phase noise in MIMO systems," *IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 347-352, Apr. 2012.

[17] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, P., and F. Capolino, "Millimeter-wave massive MIMO: the next wireless revolution?," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56-62, Sep. 2014.

[18] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Trans. Circ. Sys. I: Fundamental Theory and Applications*, vol. 47, no. 5, pp. 655-674, May 2000.

[19] F. Simoens and M. Moeneclaey, "Code-aided estimation and detection on time-varying correlated MIMO channels: A factor graph approach," *EURASIP J. Applied Signal Processing*, vol. 2006, no. 1, pp. 1-11, 2006.

[20] K. Huber, and S. Haykin, "Improved Bayesian MIMO channel tracking for wireless communications: incorporating a dynamical model," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2458-2466, Sep. 2006.

[21] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498-519, Feb. 2001.

[22] T. P. Minka, "Expectation propagation for approximate Bayesian inference," *Tech. Rep.*, Microsoft Research Cambridge, 2005, MSR-TR-2005-173 (2005).

[23] J.J. Boutros, A. Guillen i Fabregas, E. Biglieri, and G. Zemor, "Low-Density Parity-check Codes for Nonergodic Block-Fading Channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4286-4300, Sep. 2010.

[24] Y. Zhu, D. Guo, and M. L. Honig, "A message-passing approach for joint channel estimation, interference mitigation, and decoding," *IEEE Wireless Commun.*,, vol. 8, no. 12, pp. 6008-6018, Dec. 2009.

[25] K. V. Mardia, C. C. Taylor, and G. K. Subramaniam, "Protein bioinformatics and mixtures of bivariate Von Mises distributions for angular Data," *Biometrics* vol. 63, no. 2, 505-512, 2007.

[26] K. V. Mardia, J. T. Kent, G. Hughes, and C. C. Taylor, "Maximum likelihood estimation using composite likelihoods for closed exponential families," *Biometrika*, vol. 96, no. 4, pp. 975-982, 2009.

[27] H. J. Kushner, "Approximations to optimal nonlinear filters," *IEEE Trans. Automatic Control*, vol. 12, no. 5, pp. 546-556, Oct. 1967.

[28] H. J. Kushner, "Numerical approximations to optimal nonlinear filters," *in the Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovski, Eds. Oxford, 2011, ch. 28.

[29] N. Gordon, D. Salmond, and A. F. M. Smith, "Novel approach to nonlinear and non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. Eng., F,* vol. 140, pp. 107-113, 1993.

[30] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Univ. Coll. London, 1998.

[31] L. D. Darryl 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.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory,* vol. 1, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[33] M. R. Khanzadi, R. Krishnan, and T. Eriksson, "Estimation of phase noise in oscillators with colored noise sources," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2160-2163, Nov. 2013.

[34] 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.

[35] A. Tarable, G. Montorsi, S. Benedetto, and S. Chinnici, "An EM-based phase-noise estimator for MIMO systems," *IEEE Int. Conf. Commun., 2013. ICC '13.*, pp. 1-5, 8-13 Jun. 2013.

[36] A. O. Isikman, H. Mehrpouyan, A. Nasir, A. G. Amat, R. Kennedy, "Joint phase noise estimation and data detection in coded multi-inputmulti-output systems," *IET Commun.*, vol.8, no.7, pp. 981-989, May 2014.

[37] N. Noels, J. Bhatti, H. Bruneel, M. Moeneclaey, "Block-Processing soft-input soft-output demodulator for coded PSK using DCT-based phase noise estimation," *IEEE Trans. Commun.*, vol.62, no.8, pp. 2939-2950, Aug. 2014.