



HAL
open science

Approximate ML Detection for MIMO Phase Noise Channels

Richard Combes, Sheng Yang

► **To cite this version:**

Richard Combes, Sheng Yang. Approximate ML Detection for MIMO Phase Noise Channels. The International Zurich Seminar on Information and Communication, Feb 2018, Zurich, Switzerland. 10.3929/ethz-b-000245082 . hal-01730194

HAL Id: hal-01730194

<https://hal.science/hal-01730194>

Submitted on 9 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approximate ML Detection for MIMO Phase Noise Channels

Richard Combes and Sheng Yang

CentraleSupélec, L2S

91190 Gif sur Yvette, France

Email: {richard.combes, sheng.yang}@l2s.centralesupelec.fr

Abstract—We consider the MIMO communication channel impaired by phase noises at the transmitter and receiver. We focus on maximum likelihood detection for uncoded single-carrier transmission. We derive an approximation of the likelihood function, and propose the so-called *self-interference whitening* (SIW) algorithm. While the exact ML solution is computationally intractable, we construct a simulation-based lower bound on the error probability of ML detection. Numerical experiments demonstrate that SIW is, in most cases of interest, very close to optimal with moderate phase noise. Surprisingly, such near-ML performance can be achieved by applying only twice the nearest neighbor detection algorithm.

I. INTRODUCTION

We consider the signal detection problem for the following discrete-time multiple-input multiple-output (MIMO) channel

$$\mathbf{y} = \text{diag}(e^{j\theta_{t,1}}, \dots, e^{j\theta_{t,n_t}}) \mathbf{H} \text{diag}(e^{j\theta_{r,1}}, \dots, e^{j\theta_{r,n_r}}) \mathbf{x} + \mathbf{z}, \quad (1)$$

where $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$ is the channel matrix known to the receiver; $\mathbf{z} \in \mathbb{C}^{n_r \times 1}$ represents a realization of the additive noise whereas $\theta_{t,l}$ and $\theta_{r,k}$ are the phase noises at the l th transmit antenna and the k th receive antenna, respectively; the input vector $\mathbf{x} \in \mathbb{C}^{n_t \times 1}$ is assumed to be carved from a quadratic amplitude modulation (QAM). The goal is to estimate \mathbf{x} from the observation $\mathbf{y} \in \mathbb{C}^{n_r \times 1}$, with only statistical knowledge on the additive noise and the phase noises.

When phase noise is absent, the problem is well understood, and the maximum likelihood (ML) solution can be found using a nearest neighbor detection (NND) algorithm (see [1] and references therein). For instance, the sphere decoder [2] is an efficient NND with low expected complexity dimension with respect to the dimension n_t [3]. Further, there also exist approximate NND algorithms, e.g., based on lattice reduction, with near-ML performance when used for MIMO detection [4].

The presence of phase noise in (1) is a practical, long-standing problem in communication. In the seminal [5] back in the 70's, Foschini *et al.* used this model to capture the residual phase jitter at the phase-locked loop of the receiver side, and investigated both the receiver performance and the constellation design in the scalar case ($n_t = n_r = 1$). In fact, most communication systems feature phase noise due to the phase and frequency instabilities in the carrier frequency oscillators at both the transmitter and the receiver [6]. The channel (1) is a valid mathematical model when the phase

noise varies slowly as compared to the symbol duration.¹ While phase noise can be practically ignored in conventional MIMO systems, its impact becomes prominent at higher carrier frequencies since it can be shown that phase noise power increases *quadratically* with carrier frequency [6], [9]. The performance degradation due to phase noise becomes even more severe with the use of higher order modulations for which the angular separation between constellation points can be small. At medium to high SNR, phase noise dominates additive noise, becoming the capacity bottleneck [10], [11]. As for signal detection, finding the ML solution for the MIMO phase noise channel (1) is hard in general. Indeed, unlike for conventional MIMO channels, the likelihood function of the transmitted signal cannot be obtained in closed form.

In this work, we propose an efficient MIMO detection algorithm which finds an approximate ML solution in the presence of phase noise. The main contributions of this work are summarized as follows. First we derive a tractable approximation of the likelihood function of the transmitted signal. While the exact likelihood does not have a closed-form expression, the proposed approximation has a simple form and turns out to be accurate for weak to medium phase noises. Then we propose a heuristic method that finds an approximate solution by applying twice the nearest neighbor detection algorithm. The proposed algorithm, called *self-interference whitening* (SIW), has a simple geometric interpretation. Intuitively, the phase noise perturbation generates self-interference that depends on the transmitted signal through the covariance matrix. The main idea is to first estimate the covariance of the self-interference with a potentially inaccurate initial signal solution, then perform the whitening with the estimated covariance, followed by a second detection. From the optimization point of view, our algorithm can be seen as a (well-chosen) concave approximation to a non-concave objective function. Finally we assess the performance of SIW and competing algorithms in different communication scenarios. Since the error probability of ML decoding is unknown, we propose a simulation-based lower bound which we use as a benchmark. Simulation re-

¹As pointed out in [7] and the references therein, an effective discrete-time channel is usually obtained from a waveform phase noise channel after filtering. When the continuous-time phase noise varies rapidly during the symbol period, the filtered output also suffers from amplitude perturbation. See the full version of this paper [8] for further discussion.

sults show that SIW achieves near ML performance in most scenarios. In this sense, our work reveals that near optimal MIMO detection with phase noise can be done as efficiently as without phase noise.

II. ASSUMPTIONS AND PROBLEM FORMULATION

Notation: For random quantities, we use upper case letters, e.g., X , for scalars, upper case letters with bold and non-italic fonts, e.g., \mathbf{V} , for vectors, and upper case letter with bold and sans serif fonts, e.g., \mathbf{M} , for matrices. Deterministic quantities are denoted in a rather conventional way with italic letters, e.g., a scalar x , a vector \mathbf{v} , and a matrix \mathbf{M} . The Euclidean norm of a vector \mathbf{v} is denoted by $\|\mathbf{v}\|$. The transpose and conjugated transpose of \mathbf{M} are \mathbf{M}^T and \mathbf{M}^H , respectively.

We assume a MIMO channel with n_t transmit and n_r receive antennas. Let \mathbf{H} denote the channel matrix, where the (k, l) -th element of \mathbf{H} , denoted as $h_{k,l}$, represents the channel gain between the l th transmit antenna and k th receive antenna. The transmitted vector is denoted by $\mathbf{x} = [x_1, \dots, x_{n_t}]^T$, where $x_l \in \mathcal{X}, l = 1, \dots, n_t$, \mathcal{X} being typically a QAM constellation with normalized average energy, i.e., $\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} |x|^2 = 1$. The base-band output is the following random vector

$$\mathbf{Y} = \mathbf{\Lambda}_R \mathbf{H} \mathbf{\Lambda}_T \mathbf{x} + \mathbf{Z},$$

where the diagonal matrices $\mathbf{\Lambda}_R := \text{diag}(e^{j\Theta_{r,1}}, \dots, e^{j\Theta_{r,n_r}})$ and $\mathbf{\Lambda}_T := \text{diag}(e^{j\Theta_{t,1}}, \dots, e^{j\Theta_{t,n_t}})$ capture the phase perturbation at the receiver and transmitter, respectively; $\mathbf{Z} \sim \mathcal{CN}(0, \gamma^{-1} \mathbf{I})$ is the additive white Gaussian noise (AWGN) vector, where γ is the nominal signal-to-noise ratio (SNR). The phase noise $\Theta := [\Theta_{t,1} \ \dots \ \Theta_{t,n_t} \ \Theta_{r,1} \ \dots \ \Theta_{r,n_r}]^T$ is jointly Gaussian with $\Theta \sim \mathcal{N}(0, \mathbf{Q}_\theta)$ where the covariance matrix \mathbf{Q}_θ can be arbitrary. This model includes as a special case the uplink channel in which n_t is the number of single-antenna users. In such a case, the transmit phase noises are independent. We consider uncoded transmission so that each symbol x_l can take any value from \mathcal{X} with equal probability.

Further, we assume that the channel matrix can be random but is perfectly known at the receiver, whereas such knowledge at the transmitter side is irrelevant in uncoded transmission. We also define $\mathbf{H}_\Theta := \mathbf{\Lambda}_R \mathbf{H} \mathbf{\Lambda}_T$ and accordingly \mathbf{H}_θ for some realization of $\Theta = \theta$, thus, $\mathbf{H}_0 = \mathbf{H}$. Finally, we ignore the temporal correlation of the phase noise process and the channel process, and focus on the spatial aspect of the problem.

With AWGN, we have the following conditional probability density function (pdf)

$$p(\mathbf{y} | \mathbf{x}, \theta, \mathbf{H}) = \frac{\gamma^{n_r}}{\pi^{n_r}} e^{-\gamma \|\mathbf{y} - \mathbf{H}_\theta \mathbf{x}\|^2},$$

and the likelihood function by integrating over Θ

$$p(\mathbf{y} | \mathbf{x}, \mathbf{H}) = \ln \left(\mathbb{E}_\Theta \left[e^{-\gamma \|\mathbf{y} - \mathbf{H}_\Theta \mathbf{x}\|^2} \right] \right) + \ln \frac{\gamma^{n_r}}{\pi^{n_r}}.$$

The ML detector finds an input vector from the alphabet \mathcal{X}^{n_t} such that the likelihood function is maximized. In practice, it is often more convenient to use the log-likelihood function as the objective function, i.e., after removing a constant term,

$$f(\mathbf{x}, \mathbf{y}, \mathbf{H}, \gamma, \mathbf{Q}_\theta) := \ln \left(\mathbb{E}_\Theta \left[e^{-\gamma \|\mathbf{y} - \mathbf{H}_\Theta \mathbf{x}\|^2} \right] \right),$$

where the arguments γ and \mathbf{Q}_θ can be omitted whenever confusion is not likely. Thus,

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}, \mathbf{H}) := \arg \max_{\mathbf{x} \in \mathcal{X}^{n_t}} f(\mathbf{x}, \mathbf{y}, \mathbf{H}). \quad (2)$$

From (2) we see two main challenges to compute the optimal solution. First, the expectation in (2) cannot be obtained in closed form. A numerical implementation is equivalent to finding the numerical integral in $n_t + n_r$ dimensions. This can be extremely hard in high dimensions. Second, the size of the optimization space, $|\mathcal{X}|^{n_t}$, can be prohibitively large when the modulation size $|\mathcal{X}|$ and the input dimension n_t become large. In the full paper [8], we examine in more details why both of these issues are indeed challenging.

In a conventional MIMO channel, finding the ML solution is reduced to solving the following problem

$$\hat{\mathbf{x}}_{\text{ML}}^0(\mathbf{y}, \mathbf{H}) := \arg \min_{\mathbf{x} \in \mathcal{X}^{n_t}} \|\mathbf{y} - \mathbf{H}_0 \mathbf{x}\|^2, \quad (3)$$

which is also called the minimum Euclidean distance detection or nearest neighbor detection. Although the search space in (3) remains large, the expectation is gone. Furthermore, since the objective function is the Euclidean distance, efficient algorithms (e.g., sphere decoder [2] or lattice decoder [1]) exploiting the geometric structure of the problem can be applied without searching over the whole space \mathcal{X}^{n_t} . Indeed, the sphere decoder has a polynomial average complexity with respect to the input dimension n_t when the channel matrix is drawn i.i.d. from a Rayleigh distribution [3].

In practice, one may simply ignore the existence of phase noise and still apply (3) to obtain $\hat{\mathbf{x}}_{\text{ML}}^0$ which we refer to as the *naive ML* solution hereafter. While this can work relatively well when the phase noise is close to 0, it becomes highly suboptimal with stronger phase noise which is usually the case in high frequency bands with imperfect oscillators. In this paper, we provide a near ML solution by circumventing the two challenges mentioned earlier. We first propose an approximation of the likelihood function. Then we propose an algorithm to solve approximately the optimization problem (2).

III. PROPOSED SCHEME

A. Proposed Approximation of the Likelihood Function

We derive an approximation of the likelihood when the phase noise is small. Indeed, in practice, the standard deviation of the phase noise is typically smaller than 10 degrees ≈ 0.174 rad. For stronger phase noises, it is not reasonable to use QAM and the problem should be addressed differently. Consider the following approximation:

$$\begin{aligned} \mathbf{\Lambda}_R^H \mathbf{y} - \mathbf{H} \mathbf{\Lambda}_T \mathbf{x} &= [-\mathbf{H} \mathbf{D}_x \quad \mathbf{D}_y] \begin{bmatrix} e^{j\theta_t} \\ e^{-j\theta_r} \end{bmatrix} \\ &\approx (\mathbf{y} - \mathbf{H} \mathbf{x}) - j[\mathbf{H} \mathbf{D}_x \quad \mathbf{D}_y] \boldsymbol{\theta}, \end{aligned} \quad (4)$$

with $\mathbf{D}_x := \text{diag}(x_1, \dots, x_{n_t})$, $\mathbf{D}_y := \text{diag}(y_1, \dots, y_{n_r})$, and $\boldsymbol{\theta} := [\boldsymbol{\theta}_t^T \quad \boldsymbol{\theta}_r^T]^T$; (4) is from the linear approximation²

²Here we use, with a slight abuse of notation, $e^{j\boldsymbol{\theta}}$ to denote the vector obtained from the element-wise complex exponential operation. Similarly, the little- o Landau notation $o(\boldsymbol{\theta})$ is element-wise.

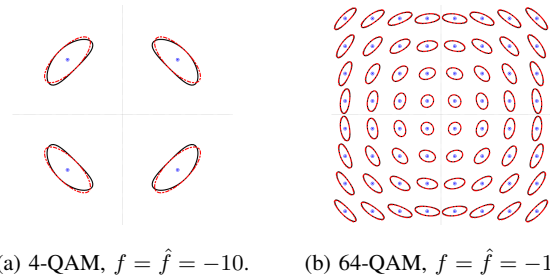


Fig. 1: Proposed approximate likelihood in the scalar case: $\gamma = 30\text{dB}$ and phase noise has standard deviation 3° .

$e^{j\theta} = 1 + j\theta + o(\theta)$. Thus the Euclidean norm has the real approximation $\|\mathbf{y} - \Lambda_{\text{R}}\mathbf{H}\Lambda_{\text{T}}\mathbf{x}\|^2 \approx \|\mathbf{A}\theta + \mathbf{b}\|^2$ where $\mathbf{A} \in \mathbb{R}^{2n_r \times (n_t + n_r)}$ and $\mathbf{b} \in \mathbb{R}^{2n_r \times 1}$ are defined as

$$\mathbf{A} := \begin{bmatrix} \Im[\mathbf{H}\mathbf{D}_x] & \Im[\mathbf{D}_y] \\ -\Re[\mathbf{H}\mathbf{D}_x] & -\Re[\mathbf{D}_y] \end{bmatrix}, \quad \mathbf{b} := \begin{bmatrix} \Re[\mathbf{y} - \mathbf{H}\mathbf{x}] \\ \Im[\mathbf{y} - \mathbf{H}\mathbf{x}] \end{bmatrix}. \quad (5)$$

The above approximation leads to the following result.

Proposition 1. *Let \mathbf{A} and \mathbf{b} be defined as in (5). Then we have the following approximation of the log-likelihood function $\ln(\mathbb{E}_{\Theta} [e^{-\gamma\|\mathbf{y} - \mathbf{H}\theta\|^2}]) \approx \hat{f}(\mathbf{x}, \mathbf{y}, \mathbf{H}, \gamma, \mathbf{Q}_\theta)$ with*

$$\hat{f}(\mathbf{x}, \mathbf{y}, \mathbf{H}) := -\gamma \mathbf{b}^T \mathbf{W}_{\mathbf{x}}^{-1} \mathbf{b} - \frac{1}{2} \ln \det(\mathbf{W}_{\mathbf{x}}), \quad \text{and} \quad (6)$$

$$\mathbf{W}_{\mathbf{x}} := \mathbf{I} + 2\gamma \mathbf{A} \mathbf{Q}_\theta \mathbf{A}^T. \quad (7)$$

Hence, the proposed approximate ML (aML) solution is

$$\hat{\mathbf{x}}_{\text{aML}}(\mathbf{y}, \mathbf{H}) := \arg \min_{\mathbf{x} \in \mathcal{X}^{n_t}} \left\{ \gamma \mathbf{b}^T \mathbf{W}_{\mathbf{x}}^{-1} \mathbf{b} + \frac{1}{2} \ln \det(\mathbf{W}_{\mathbf{x}}) \right\}. \quad (8)$$

Proof. The proof is straightforward after applying the above approximation. Details can be found in the full paper [8]. \square

In Fig. (1), we illustrate the proposed approximation for 4- and 64-QAM, we plot for each constellation point a level set of the likelihood function with respect to “ \mathbf{y} ” in solid line. The level sets of the approximated likelihood function are plotted similarly in dashed line. While the likelihood function is evaluated using numerical integration, the approximation is in closed form given by (6). In this figure, we observe that the approximation is quite accurate, especially for signal points with smaller amplitude. Further, the resemblance of the level sets for the approximate likelihood to ellipsoids suggests that the main contribution in the right hand side of (6) comes from the first term $-\gamma \mathbf{b}^T \mathbf{W}_{\mathbf{x}}^{-1} \mathbf{b}$. We shall exploit this feature later on to construct the proposed algorithm.

While the proposed approximation simplifies significantly the objective function, the optimization problem (8) remains hard when the search space is large. For instance, with 64-QAM and 4×4 MIMO, the number of points in \mathcal{X}^{n_t} is more than 10^7 ! Therefore, we need further simplification by exploiting the structure of the problem.

B. The Self-Interference Whitening Algorithm

The difficulty of the optimization (8) is mainly due to the presence of the matrix $\mathbf{W}_{\mathbf{x}}$ that depends on \mathbf{x} . Let us first assume that the $\mathbf{W}_{\mathbf{x}}$ corresponding to the optimal solution $\hat{\mathbf{x}}_{\text{aML}}$ were somehow known, and is denoted by $\mathbf{W}_{\hat{\mathbf{x}}}$. Then the optimization problem (8) would be equivalent to

$$\begin{aligned} \hat{\mathbf{x}}_{\text{aML}}(\mathbf{y}, \mathbf{H}) &= \arg \min_{\mathbf{x} \in \mathcal{X}^{n_t}} \left\{ \gamma \mathbf{b}^T \mathbf{W}_{\hat{\mathbf{x}}}^{-1} \mathbf{b} + \frac{1}{2} \ln \det(\mathbf{W}_{\hat{\mathbf{x}}}) \right\} \\ &= \arg \min_{\mathbf{x} \in \mathcal{X}^{n_t}} \mathbf{b}^T \mathbf{W}_{\hat{\mathbf{x}}}^{-1} \mathbf{b} \\ &= \arg \min_{\mathbf{x} \in \mathcal{X}^{n_t}} \left\| \mathbf{W}_{\hat{\mathbf{x}}}^{-\frac{1}{2}} (\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}) \right\|^2, \end{aligned} \quad (9)$$

where $\mathbf{W}_{\hat{\mathbf{x}}}^{-\frac{1}{2}}$ is any matrix such that $(\mathbf{W}_{\hat{\mathbf{x}}}^{-\frac{1}{2}})^H \mathbf{W}_{\hat{\mathbf{x}}}^{-\frac{1}{2}} = \mathbf{W}_{\hat{\mathbf{x}}}^{-1}$;

$$\tilde{\mathbf{x}} := \begin{bmatrix} \Re[\mathbf{x}] \\ \Im[\mathbf{x}] \end{bmatrix}, \quad \tilde{\mathbf{y}} := \begin{bmatrix} \Re[\mathbf{y}] \\ \Im[\mathbf{y}] \end{bmatrix}, \quad \tilde{\mathbf{H}} := \begin{bmatrix} \Re[\mathbf{H}] & -\Im[\mathbf{H}] \\ \Im[\mathbf{H}] & \Re[\mathbf{H}] \end{bmatrix}. \quad (10)$$

Note that for a given $\mathbf{W}_{\hat{\mathbf{x}}}$, (9) can be solved with any NND algorithm. Unfortunately, without knowing the optimal solution $\hat{\mathbf{x}}_{\text{aML}}$, the exact $\mathbf{W}_{\hat{\mathbf{x}}}$ cannot be found. Therefore, the idea is to first estimate the matrix $\mathbf{W}_{\hat{\mathbf{x}}}$ with some suboptimal solution $\hat{\mathbf{x}}$, and then solve the optimization problem (9) with a NND. We call this two-step procedure *self-interference whitening* (SIW). For instance, we can use the naive ML solution $\hat{\mathbf{x}}_{\text{ML}}^0$ as the initial estimate to obtain $\mathbf{W}_{\hat{\mathbf{x}}}$, and have $\hat{\mathbf{x}}'_{\text{aML}}(\mathbf{y}, \mathbf{H}) = \arg \min_{\mathbf{x} \in \mathcal{X}^{n_t}} \left\| \mathbf{W}_{\hat{\mathbf{x}}_{\text{ML}}^0}^{-\frac{1}{2}} \tilde{\mathbf{y}} - \mathbf{W}_{\hat{\mathbf{x}}_{\text{ML}}^0}^{-\frac{1}{2}} \tilde{\mathbf{H}}\tilde{\mathbf{x}} \right\|^2$.

Remark 1. *The intuition behind the SIW scheme is as follows. From the definition of $\mathbf{W}_{\mathbf{x}}$ in (7) and \mathbf{A} in (5), we see that $\mathbf{W}_{\mathbf{x}}$ depends on \mathbf{x} only through $\mathbf{H}\mathbf{D}_x$. First, the column space of $\mathbf{H}\mathbf{D}_x$ does not vary with \mathbf{x} since \mathbf{D}_x is diagonal. Second, a small perturbation of \mathbf{x} does not perturb $\mathbf{W}_{\mathbf{x}}$ too much. Since the naive ML point $\hat{\mathbf{x}}_{\text{ML}}^0$ is close to the actual point \mathbf{x} in the column space of \mathbf{H} , it provides an accurate estimate of $\mathbf{W}_{\mathbf{x}}$. This can also be observed on Fig. (1b), where we see that the ellipsoid-like dashed lines have similar sizes and orientations for constellation points that are close to each other.*

Remark 2. *Another possible initial estimate is the naive linear minimum mean square error (LMMSE) solution. As the naive ML, the naive LMMSE ignores the phase noise and returns*

$$\hat{\mathbf{x}}_{\text{LMMSE}}^0(\mathbf{y}, \mathbf{H}) := \arg \min_{\mathbf{x} \in \mathcal{X}^{n_t}} \left\| \mathbf{H}^H (\gamma^{-1} \mathbf{I} + \mathbf{H}\mathbf{H}^H)^{-1} \mathbf{y} - \mathbf{x} \right\|^2. \quad (11)$$

The SIW algorithm is described in Algorithm 1. Here, the complex function $\text{NND}(\mathbf{y}, \mathbf{H}, \mathcal{X})$ finds among the points from the alphabet \mathcal{X} the closest one to \mathbf{y} in the column space of \mathbf{H} ; the function $\text{realNND}(\tilde{\mathbf{y}}, \tilde{\mathbf{H}}, \tilde{\mathcal{X}})$ is the real counterpart of NND. The function “complex($\tilde{\mathbf{x}}'$)” embeds the real vector $\tilde{\mathbf{x}}'$ to the complex space by taking the upper half as the real part and the lower half as the imaginary part. The SIW outputs the newly obtained point only if it has a higher approximate likelihood value than the naive ML point does. An example of the scalar case using 256-QAM is shown in Fig. (2). The transmitted point is x and the received point is y . The solid line is the level set of the likelihood function. If

Algorithm 1 Self-interference whitening

Input: $\mathbf{y}, \mathbf{H}, \gamma, \mathbf{Q}_\theta$
 Find $\hat{\mathbf{x}}_{\text{LMMSE}}^0$ from (11)
 Find $\hat{\mathbf{x}}_{\text{ML}}^0 \leftarrow \text{NND}(\mathbf{y}, \mathbf{H}, \mathcal{X})$
 if $\hat{f}(\hat{\mathbf{x}}_{\text{LMMSE}}^0, \mathbf{y}, \mathbf{H}, \gamma, \mathbf{Q}_\theta) > \hat{f}(\hat{\mathbf{x}}_{\text{ML}}^0, \mathbf{y}, \mathbf{H}, \gamma, \mathbf{Q}_\theta)$ then
 $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}}_{\text{LMMSE}}^0$
 else
 $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}}_{\text{ML}}^0$
 end if
 Generate $\mathbf{W}_{\hat{\mathbf{x}}}$ from $\hat{\mathbf{x}}$ using (5) and (7)
 Find $\mathbf{W}_{\hat{\mathbf{x}}}^{\frac{1}{2}}$ using the Cholesky decomposition
 Generate $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{H}}$ according to (10)
 $\hat{\mathbf{x}}' \leftarrow \text{realNND}(\mathbf{W}_{\hat{\mathbf{x}}}^{-\frac{1}{2}} \tilde{\mathbf{y}}, \mathbf{W}_{\hat{\mathbf{x}}}^{-\frac{1}{2}} \tilde{\mathbf{H}} \hat{\mathbf{x}}, \tilde{\mathcal{X}})$
 $\hat{\mathbf{x}}' \leftarrow \text{complex}(\hat{\mathbf{x}}')$
 if $\hat{f}(\hat{\mathbf{x}}', \mathbf{y}, \mathbf{H}, \gamma, \mathbf{Q}_\theta) > \hat{f}(\hat{\mathbf{x}}, \mathbf{y}, \mathbf{H}, \gamma, \mathbf{Q}_\theta)$ then
 $\hat{\mathbf{x}}'_{\text{aML}} \leftarrow \hat{\mathbf{x}}'$
 else
 $\hat{\mathbf{x}}'_{\text{aML}} \leftarrow \hat{\mathbf{x}}$
 end if
 Output: $\hat{\mathbf{x}}'_{\text{aML}}$

the likelihood function were computed for each point in the constellation (this is computationally hard), one would recover x from y successfully. While Euclidean detection outputs the wrong point \hat{x} , SIW can “correct” the error by first estimating the unknown matrix \mathbf{W}_x , and then computing the matrix $\mathbf{W}_{\hat{x}}$ which is represented by the red dashed ellipse around \hat{x} . The estimate $\mathbf{W}_{\hat{x}}$ is very close to the correct value \mathbf{W}_x , given by the actual x (blue dashed line). Then, SIW searches for the closest constellation point to y in the coordinate system generated by $\mathbf{W}_{\hat{x}}$, so that x is recovered successfully. Also, computationally efficient NND algorithms can be used to perform the search.

The complexity of the SIW algorithm is essentially twice that of the NND algorithm used, since the other operations including the LMMSE detection have at most cubic complexity with respect to the dimension of the channel. The complexity of the NND algorithm depends directly on the conditioning of the given matrix. If the columns are close to orthogonal, then channel inversion is almost optimal. However, in the worse case, when the matrix is ill-conditioned, the NND algorithm can be slow and its complexity is exponential in the problem dimension. There exist approximate NND algorithms, e.g., based on lattice reduction, that can achieve near optimal performance with much lower complexity.

IV. NUMERICAL EXPERIMENTS

We now compare the performance³ of SIW to ML and other baseline schemes: i) the naive LMMSE (11), ii) the naive ML (3), and iii) the selection between the two where the receiver outputs the one whose approximate likelihood value is higher. We derive a lower bound on the performance of

³Our performance metric is the vector detection error rate: detection is considered successful only when all the symbols in \mathbf{x} are recovered correctly.

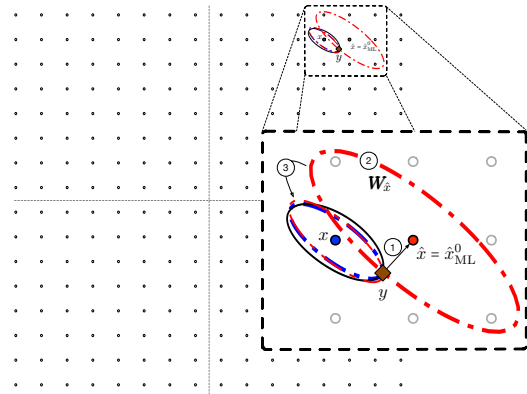


Fig. 2: SIW in the scalar case: 256-QAM, PN 2°. Dashed lines are the ellipses defined by $\mathbf{W}_{\hat{x}}$ (in red) and \mathbf{W}_x (in blue).

ML detection since it is hard to implement. A scheme minimizing (6) admits the following performance lower bound:

$$P_e^{\text{aML}} \geq \mathbb{P} \left\{ \hat{f}(\mathbf{X}, \mathbf{Y}, \mathbf{H}) < \max_{\mathbf{x} \in \mathcal{X}^{n_t}} \hat{f}(\mathbf{x}, \mathbf{Y}, \mathbf{H}) \right\} \quad (12)$$

$$\geq \mathbb{P} \left\{ \hat{f}(\mathbf{X}, \mathbf{Y}, \mathbf{H}) < \max_{\mathbf{x} \in \mathcal{L} \subseteq \mathcal{X}^{n_t}} \hat{f}(\mathbf{x}, \mathbf{Y}, \mathbf{H}) \right\}, \quad (13)$$

where (12) holds by definition and (13) by monotonicity. While (13) holds for all \mathcal{L} , one has equality if \mathcal{L} contains *all* the points in \mathcal{X}^{n_t} that have a higher approximate likelihood value than \mathbf{X} does. Here, we consider a large set around \mathbf{X} to compute the lower bound (13), but do not study its tightness. Similarly, for ML detection, we have

$$P_e^{\text{ML}} \geq \mathbb{P} \left\{ f(\mathbf{X}, \mathbf{Y}, \mathbf{H}) < \max_{\mathbf{x} \in \mathcal{X}^{n_t}} f(\mathbf{x}, \mathbf{Y}, \mathbf{H}) \right\} \quad (14)$$

$$\geq \mathbb{P} \{ f(\mathbf{X}, \mathbf{Y}, \mathbf{H}) < f(\mathbf{X}', \mathbf{Y}, \mathbf{H}) \}, \quad \forall \mathbf{X}' \in \mathcal{X}^{n_t} \quad (15)$$

$$= \mathbb{P} \{ \mathbf{X} \neq \mathbf{X}', f(\mathbf{X}, \mathbf{Y}, \mathbf{H}) < f(\mathbf{X}', \mathbf{Y}, \mathbf{H}) \}, \quad (16)$$

where (14) holds by definition and (16) holds since $\mathbf{X} \neq \mathbf{X}'$ is a consequence of $f(\mathbf{X}, \mathbf{Y}, \mathbf{H}) < f(\mathbf{X}', \mathbf{Y}, \mathbf{H})$. Also, (15) holds for any $\mathbf{X}' \in \mathcal{X}^{n_t}$, with equality if \mathbf{X}' is the exact ML solution. Since the ML solution is unknown, one may use any suboptimal solution instead and to obtain a lower bound. Indeed (16) is much easier to evaluate than (14), as the latter requires to minimize over \mathcal{X}^{n_t} . Intuitively, if \mathbf{X}' is a near ML solution, then the lower bound should be tight enough. We need to perform twice the numerical integration only when $\mathbf{x}' \neq \mathbf{x}$. If $\mathbf{x}' \neq \mathbf{x}$ with small probability, evaluating (16) can be done quickly.

In Fig. (3a) we consider point-to-point Rayleigh fading single-antenna, i.e., single-input single-output (SISO), channels. We consider 1024-QAM with phase noise of standard deviation 1° at both the transmitter and receiver sides. Here the naive ML scheme is in fact a simple threshold detection for the real and imaginary parts. First, we see that ignoring the existence of phase noise incurs a significant performance loss. Second, if exhaustive search is done with the proposed likelihood approximation, then it achieves the ML performance,

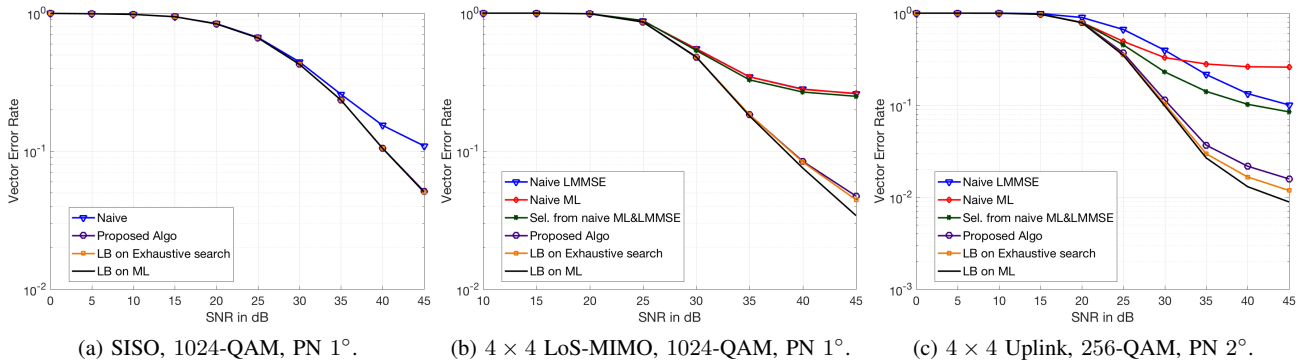


Fig. 3: Simulation results for different communication scenarios.

since the simulation-based lower bound overlaps with the curve with exhaustive search. This confirms the accuracy of the approximation (6) in the SISO case. Even more remarkably, the SIW algorithm almost achieves the ML performance without exhaustive search.

In Fig. (3b) we consider point-to-point line-of-sight (LoS) MIMO, commonly deployed as microwave backhaul links [10], [12], [13]. We assume that the channel is constant over time but each antenna is driven by its own oscillator. This is the worst-case assumption, and motivated in practice by the fact that the communication distance is large and thus the distance between antenna elements is increased so that the channel matrix is well conditioned [12], [13]. We use the model of [13] with two transmit and two receive antennas each with dual polarizations, effectively a 4×4 MIMO channel. The optimal distance between the antenna elements at each side can be derived as a function of the communication distance [12], for which the channel matrix is unitary. As above we consider 1024-QAM with phase noise of standard deviation 1° . As in the SISO case, phase noise mitigation substantially improves the performance and the proposed likelihood approximation remains accurate as shown by the comparison between the exhaustive search (8) and the lower bound on ML detection.

In Fig. (3c) we consider the uplink cellular communication channel with four single-antenna users and one multi-antenna base station receiver. We assume i.i.d. phase noises at the users' side with standard deviation 2° and no phase noise at the receiver side. This is a reasonable assumption since the oscillators at the base station are usually of higher quality than those used by mobile devices. We assume i.i.d. Rayleigh fading. Unlike in the previous scenarios, the naive ML is (surprisingly) dominated by the naive LMMSE at high SNR. Indeed, without receiver phase noise, inverting the channel yields spatial parallel channels. Although this incurs a power loss in general, phase noises across the parallel sub-channels are independent, so the demodulation only suffers from a scalar self-interference. On the other hand, naive ML suffers from the aggregated perturbation from all the phase noises. So naive LMMSE beats naive ML detection at high SNRs

where phase noise dominates the additive noise. If both the transmitter and receiver have comparable phase noises, this does not occur, as channel inversion amplifies the receiver phase noises. The gain of SIW over the other schemes is clear.

V. CONCLUSIONS

We have studied the ML detection problem for uncoded MIMO phase noise channels, and proposed an approximation of the likelihood function that has been shown to be accurate in the regimes of practical interest. More importantly, using the geometric interpretation of the approximate likelihood function, we have designed SIW, an efficient approximate algorithm requiring only two nearest neighbor detections.

REFERENCES

- [1] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, Aug. 2002.
- [2] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channel," *IEEE Trans. Inf. Theory*, vol. 45, no. 10, Jul. 1999.
- [3] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Sig. Proc.*, vol. 53, no. 8, Aug. 2005.
- [4] J. Jaldén and P. Elia, "DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs, and system models," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, Oct. 2010.
- [5] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "On the selection of a two-dimensional signal constellation in the presence of phase jitter and gaussian noise," *Bell Labs Tech. J.*, vol. 52, no. 6, 1973.
- [6] A. Hajimiri and T. H. Lee, "A general theory of phase noise in electrical oscillators," *IEEE J. of Solid-State Circuits*, vol. 33, no. 2, Feb. 1998.
- [7] H. Ghozlan and G. Kramer, "Models and information rates for Wiener phase noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, Apr. 2017.
- [8] R. Combes and S. Yang, "An approximate ML detector for MIMO channels corrupted by phase noise," to appear in *IEEE Trans. Commun.*
- [9] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterization," *IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 5, May 2000.
- [10] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise," *IEEE Trans. Commun.*, vol. 62, no. 3, Mar. 2014.
- [11] S. Yang and S. Shamai (Shitz), "On the multiplexing gain of discrete-time MIMO phase noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, Apr. 2017.
- [12] F. Bøhagen, P. Orten, and G. E. Øien, "Design of optimal high-rank line-of-sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, Apr. 2007.
- [13] P. Ferrand and S. Yang, "Blind precoding in line-of-sight MIMO channels," in *IEEE SPAWC*, Edinburgh, UK, 2016.