On the Capacity of the Wiener Phase Noise Channel: Bounds and Capacity Achieving Distributions

Article *in* IEEE Transactions on Communications · March 2015

DOI: 10.1109/TCOMM.2015.2465389 · Source: arXiv

CITATIONS
5

READS
40

4 authors:

M. Reza Khanzadi
Chalmers University of Technology
25 PUBLICATIONS **260** CITATIONS
SEE PROFILE

Rajet Krishnan
Huawei R & D Sweden, CommSig AB
26 PUBLICATIONS **354** CITATIONS
SEE PROFILE

Johan Söder
Ericsson
1 PUBLICATION **5** CITATIONS
SEE PROFILE

Thomas Eriksson
Chalmers University of Technology
246 PUBLICATIONS **3,403** CITATIONS
SEE PROFILE

On the Capacity of the Wiener Phase Noise Channel: Bounds and Capacity Achieving Distributions

M. Reza Khanzadi, *Student Member, IEEE*, Rajet Krishnan, *Student Member, IEEE*,
Johan Söder, Thomas Eriksson

Abstract

In this paper, the capacity of the additive white Gaussian noise (AWGN) channel, affected by time-varying Wiener phase noise is investigated. Tight upper and lower bounds on the capacity of this channel are developed. The upper bound is obtained by using the duality approach, and considering a specific distribution over the output of the channel. In order to lower-bound the capacity, first a family of capacity-achieving input distributions is found by solving a functional optimization of the channel mutual information. Then, lower bounds on the capacity are obtained by drawing samples from the proposed distributions through Monte-Carlo simulations. The proposed capacity-achieving input distributions are circularly symmetric, non-Gaussian, and the input amplitudes are correlated over time. The evaluated capacity bounds are tight for a wide range of signal-to-noise-ratio (SNR) values, and thus they can be used to quantify the capacity. Specifically, the bounds follow the well-known AWGN capacity curve at low SNR, while at high SNR, they coincide with the high-SNR capacity result available in the literature for the phase noise channel.

M. Reza Khanzadi is with the Department of Signals and Systems, and also Department of Microtechnology and Nanoscience, Chalmers University of Technology, 41296 Gothenburg, Sweden. (email: khanzadi@chalmers.se.)

Rajet Krishnan, and Thomas Eriksson are with the Department of Signals and Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden. (email: rajet,thomase@chalmers.se.)

Johan Söder is with Ericsson Research, Stockholm, Sweden. (email: johan.soder@ericsson.com)

The simulations were performed on resources at Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC).

I. Introduction

Phase noise due to frequency instabilities of radio frequency oscillators is a limiting factor in high data rate digital communication systems (e.g., see [1]–[7] and references therein). Phase noise severely impacts the performance of systems that employ dense signal constellations [8], [9]. Moreover, the effect of phase noise is more pronounced in high carrier frequency systems, e.g., E-band (60-80 GHz), mainly due to the high levels of phase noise in oscillators designed for such frequencies [10]–[13].

The Shannon capacity of the system can be studied in order to investigate the effect of phase noise on the throughput. For stationary phase noise channels, Lapidoth [14] derived an asymptotic capacity expression, that is valid at high SNR. Capacity of the noncoherent channel, where the transmitted signal is affected by uniformly distributed phase noise, has been studied in [15]–[17]. In [17], Katz and Shamai derived upper and lower bounds on the capacity of the noncoherent phase noise channel for non-asymptotic SNR regimes. They showed that the capacity-achieving distribution of the noncoherent memoryless channel is discrete with an infinite number of mass points. In [15], [18], the capacity bounds in [17] have been extended to the block memoryless phase noise channel, where the phase noise was modeled as a constant over a number of consecutive symbols. It was shown in [15] that the capacity-achieving input distribution of the block memoryless phase noise channel is not Gaussian (unlike the additive white Gaussian noise channel). In [16], the constrained capacity of M-ary phase-shift keying over a noncoherent phase noise channel has been investigated. Capacity of partially coherent channels, where the phase noise is estimated at the receiver, and the signal is affected by the residual phase noise estimation errors, has been studied in [19]. Effects of using multisampling receivers on the achievable information rate of the phase noise channel has been recently investigated in [20]–[22].

There has been limited number of studies on characterizing the capacity of channels affected by phase noise with memory (e.g., [14], [20], [23], [24]). The Wiener phase noise channel that models many practical scenarios belongs to this family of channels. In [14], Lapidoth characterized the capacity of the Wiener phase noise channel at high SNR. It was shown in [14] that circularly symmetric input alphabets with Gamma-distributed amplitudes can achieve the capacity of the stationary phase noise channel (with or without memory) at high SNR. Capacity

results of [14] have been recently extended to multi-antenna systems in [23], [25]. However, the capacity-achieving input distribution of the Wiener phase noise channel and the closed-form capacity of this channel, valid for all SNR values, have not been derived yet.

Contributions

In this paper, we derive tight upper and lower bounds on the capacity of the additive white Gaussian noise (AWGN) channel affected by Wiener phase noise when the channel input is subject to an average-power constraint. The upper bound on the capacity is found by using the duality approach, and considering a specific distribution over the output of the channel. We determine a family of input distributions that result in a tight lower bound on the capacity. We show that the capacity-achieving input distribution is circularly symmetric but non-Gaussian. We also show that unlike for memoryless channels, the capacity-achieving input alphabets are correlated over time. Lower bounds on the capacity are obtained by numerical calculation of the information rates, achievable by samples generated from the proposed input distributions through Monte-Carlo simulations. The developed upper and lower bounds are tight for a wide range of SNR values. This helps to more accurately quantify the capacity of the phase noise channel compared to previously available results in the literature (e.g., [14]).

Organization of the Paper

The paper is organized as follows. In Section II, the system model and the corresponding amplitude-phase channel are introduced. Using the amplitude-phase channel, the mutual information between the input and the output is examined in Section III. In Section IV, a capacity upper bound is derived. In Section V, we obtain the closed-form expression for a family of capacity-achieving distributions. Finally, in Section VI, the proposed lower and upper bounds are compared against each other and also the results available in the literature.

Notation

With $\mathcal{N}(0,\sigma^2)$ and $\mathcal{CN}(0,\sigma^2)$, we denote the probability distribution of a real Gaussian random variable, and of a circularly symmetric complex Gaussian random variable with zero mean and variance σ^2. The uniform distribution over the interval $[0,2\pi)$ is denoted as $\mathcal{U}(0,2\pi)$. We use $|\cdot|$ to denote the absolute value of scalars, and determinant of matrices. The Euclidean norm of

vectors is denoted by $||\cdot||$. Finally, $\mathcal{D}(\cdot||\cdot)$ denotes the relative entropy between two probability distributions.

II. System Model

A. The channel

The input-output relation of the Wiener phase noise channel can be written as [2]

$$y_k = x_k e^{j\phi_k} + w_k, \tag{1}$$

where x_k is the transmitted symbol, and w_k is circular symmetric AWGN independently distributed from $\mathcal{CN}(0, 2\sigma_{\mathrm{W}}^2)$. The process, ϕ_k, is the discrete-time Wiener phase noise

$$\phi_k = \phi_{k-1} + \Delta_k, \qquad \Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2). \tag{2}$$

This discrete-time process corresponds to a sampled version of a continuous-time Brownian motion process with uncorrelated increments.[1] Samples are taken every T_{s} seconds, the transmission symbol interval.[2] The continuous time process of the corresponding oscillator has a Lorentzian spectrum [3], [27]. This spectrum is fully characterized by a single parameter; the 3dB single-sided bandwidth, $f_{3\mathrm{dB}}$, which depends on central frequency and design technology of the oscillator [28]. The innovation variance for the discrete phase noise process is $\sigma_\Delta^2 = 4\pi f_{3\mathrm{dB}} T_{\mathrm{s}}$.[3]

B. Amplitude and phase input-output relations

The input x_k, to the channel (1) and the output y_k are complex numbers, and can be represented in polar form as $x_k = R_k e^{j\Theta_k}$, and $y_k = r_k e^{j\theta_k}$. In this notation, R_k and Θ_k denote the amplitude and the phase of the transmitted symbol x_k, while r_k and θ_k denote the amplitude and the phase

[1]For discussions on the limitations of the Wiener phase noise model see [3], [6], [26].

[2]Note that the system model (1) is derived under the assumption that the continuous-time phase noise process remains constant over the duration of the symbol time. This assumption allows us to obtain a discrete-time equivalent channel model by sampling at Nyquist rate. As shown recently in [20]–[22], by dropping this assumption one may obtain different high-SNR capacity characterization.

[3]Note that the innovation variance can equivalently be found directly from the spectrum of the phase noise process [28].

of the received sample y_k, respectively. The input-output relations between the transmitted and received amplitude and phase are

$$r_k = \sqrt{(R_k + w_{k,\parallel})^2 + w_{k,\perp}^2} \tag{3}$$

$$\theta_k = \Theta_k + N_k + \phi_k, \tag{4}$$

where

$$N_k = \arctan\left(\frac{w_{k,\perp}}{R_k + w_{k,\parallel}}\right), \tag{5}$$

and $w_{k,\parallel}$ and $w_{k,\perp}$ denote the parts of w_k that are parallel (in-phase) and orthogonal to the transmitted signal, respectively. In the rest of the paper, we investigate the capacity of the phase-noise channel (1) by considering the equivalent amplitude and phase channels stated in (3) and (4).

C. Definition of Capacity

The capacity of the phase-noise channel (1) is given by [29]

$$C(\mathsf{SNR}) = \lim_{n\to\infty} \sup_{f(\mathbf{x})} \frac{1}{n} I(\mathbf{x};\mathbf{y}) \tag{6}$$

$$= \lim_{n\to\infty} \sup_{f(\mathbf{R},\boldsymbol{\Theta})} \frac{1}{n} I(\mathbf{r},\boldsymbol{\theta};\mathbf{R},\boldsymbol{\Theta}), \tag{7}$$

where $\mathbf{x} = \{x_k\}_{k=1}^n$, $\mathbf{y} = \{y_k\}_{k=1}^n$, $\mathbf{r} = \{r_k\}_{k=1}^n$, $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^n$, and $\mathbf{R} = \{R_k\}_{k=1}^n$, $\boldsymbol{\Theta} = \{\Theta_k\}_{k=1}^n$. The supremum in (6) and (7) is computed over all probability distributions on the input that satisfies

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[|x_k|^2\right] = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[R_k^2\right] \leq E_{\mathrm{s}}, \tag{8}$$

where E_{s} is the maximum average power. The SNR is defined as $\mathsf{SNR} = E_{\mathrm{s}}/2\sigma_{\mathrm{W}}^2$ throughout the paper.

III. Mutual Information

The mutual information on the right-hand side (RHS) of (7) is written as

$$\begin{aligned}
&\frac{1}{n} I(\mathbf{r}, \boldsymbol{\theta}; \mathbf{R}, \boldsymbol{\Theta}) && (9)\\
&\quad = \frac{1}{n} \left(h(\mathbf{r}, \boldsymbol{\theta}) - h(\mathbf{r}, \boldsymbol{\theta} | \mathbf{R}, \boldsymbol{\Theta}) \right) && (10)\\
&\quad = \frac{1}{n} \left(h(\mathbf{r}) + h(\boldsymbol{\theta}|\mathbf{r}) - h(\mathbf{r}|\mathbf{R}, \boldsymbol{\Theta}) - h(\boldsymbol{\theta}|\mathbf{r}, \mathbf{R}, \boldsymbol{\Theta}) \right) && (11)\\
&\quad = \frac{1}{n} \left(h(\mathbf{r}) + h(\boldsymbol{\theta}|\mathbf{r}) - h(\mathbf{r}|\mathbf{R}) - h(\boldsymbol{\theta}|\mathbf{r}, \mathbf{R}, \boldsymbol{\Theta}) \right) && (12)\\
&\quad = \frac{1}{n} \left(I(\mathbf{r}; \mathbf{R}) + h(\boldsymbol{\theta}|\mathbf{r}) - h(\boldsymbol{\Theta} + \mathbf{N} + \boldsymbol{\phi}|\mathbf{r}, \boldsymbol{\Theta}, \mathbf{R}) \right) && (13)\\
&\quad = \frac{1}{n} \left(I(\mathbf{r}; \mathbf{R}) + h(\boldsymbol{\theta}|\mathbf{r}) - h(\mathbf{N} + \boldsymbol{\phi}|\mathbf{r}, \mathbf{R}) \right), && (14)
\end{aligned}$$

where $\boldsymbol{\phi} = \{\phi_k\}_{k=1}^n$, and $\mathbf{N} = \{N_k\}_{k=1}^n$. In (11) the chain rule for entropy is used, and (12) follows because $\mathbf{r}$ and $\boldsymbol{\Theta}$ are independent (see (3)). In (13) and (14), we used the definition of θ_k, given in (4).

We present the following lemma pertaining to the capacity-achieving input distribution, which will be used throughout the paper.

Lemma 1: The capacity-achieving input of the channel (1) is circularly symmetric, i.e., $\{\Theta_k\}_{k=1}^n$ are independently and identically distributed from $\mathcal{U}(0, 2\pi)$ and are independent of $\mathbf{R}$.

Proof: The proof directly follows that of [30, Prop. 7], where it is shown that the capacity-achieving input of the fading channel with memory is circularly symmetric. ■

Based on (4) and the result of Lemma 1, it can be deduced that the output phase is also uniformly distributed and, hence, $h(\boldsymbol{\theta}|\mathbf{r}) = n \log_2 2\pi$. Thus, (14) can be rewritten as

$$\frac{1}{n} I(\mathbf{r}, \boldsymbol{\theta}; \mathbf{R}, \boldsymbol{\Theta}) = \frac{1}{n} \left(I(\mathbf{r}; \mathbf{R}) - h(\mathbf{N} + \boldsymbol{\phi}|\mathbf{r}, \mathbf{R}) \right) + \log_2 2\pi. \quad (15)$$

Next, we use the definition of the phase channel (4), and rewrite the second term on the RHS

of (15) as

$$\frac{1}{n}h(\mathbf{N}+\boldsymbol{\phi}|\mathbf{r},\mathbf{R}) \tag{16}$$

$$= \frac{1}{n}h(\underline{D}(\mathbf{N}+\boldsymbol{\phi})|\mathbf{r},\mathbf{R}) - \log_2|\underline{D}| \tag{17}$$

$$= \frac{1}{n}h(N_n - N_{n-1} + \Delta_n, \ldots, N_2 - N_1 + \Delta_2, N_1 + \Delta_1|\mathbf{r},\mathbf{R}) \tag{18}$$

$$= \frac{1}{n}h(\{N_k - N_{k-1} + \Delta_k\}_{k=2}^{n}, N_1 + \Delta_1|\mathbf{r},\mathbf{R}) \tag{19}$$

$$\triangleq \frac{1}{n}h\left(\{a_k\}_{k=1}^{n}|\mathbf{r},\mathbf{R}\right) \tag{20}$$

$$= h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r},\mathbf{R}\right), \tag{21}$$

where

$$a_k \triangleq \begin{cases} N_1 + \Delta_1 & \text{if } k = 1 \\ N_k - N_{k-1} + \Delta_k & \text{if } k > 1 \end{cases}. \tag{22}$$

In (17), we choose $\underline{D}$ to be the *difference matrix* defined as

$$\underline{D} = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \ldots & \ldots & 0 & 1 \end{bmatrix}, \tag{23}$$

where $|\underline{D}| = 1$. The equality in (17) follows from [29, Eq. 8.71], and in (18), we used that $\log_2(|\underline{D}|) = 0$. The noise vector $\mathbf{N}+\boldsymbol{\phi}$ is rearranged (17) as the difference between the consecutive noise samples in order to resolve the infinite memory of the phase noise process ϕ_k. In (21), the joint entropy (20) is replaced with an equivalent expression for the differential-entropy rate [29, p. 74]. By substituting (21) in (15), we obtain

$$\frac{1}{n}I(\mathbf{r},\boldsymbol{\theta};\mathbf{R},\boldsymbol{\Theta}) = \frac{1}{n}I(\mathbf{r};\mathbf{R}) - h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r},\mathbf{R}\right) + \log_2 2\pi. \tag{24}$$

In order to find upper and lower bounds on the capacity, we need to evaluate the two first terms on the RHS of (24).

IV. Capacity Upper Bound

In this section, an upper bound on the capacity of the phase-noise channel (1) is derived. We first find a lower bound for the entropy term on the RHS of (24). To do so, we condition on the complete knowledge of the noise sample N_{n-1} as follows

$$h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r},\mathbf{R}\right) \geq h\left(a_n|\{a_k\}_{k=1}^{n-1},N_{n-1},\mathbf{r},\mathbf{R}\right) \tag{25}$$

$$= h\left(a_n|N_{n-1},\mathbf{r},\mathbf{R}\right) \tag{26}$$

$$= h\left(N_n+\Delta_n|r_n,R_n\right). \tag{27}$$

Here, in (25), we used that conditioning reduces entropy. Equality in (26) holds because a_n is conditionally independent of $\{a_k\}_{k=1}^{n-1}$ given N_{n-1} (see (22)). Finally in (27), we removed all the other terms that are conditionally independent of $N_n+\Delta_n$.

We next upper-bound the mutual information on the RHS of (24) by using the *duality approach* [31, Th. 5.1]. Let $f(\mathbf{r}|\mathbf{R})$ denote the conditional probability of $\mathbf{r}$ given $\mathbf{R}$, and $f(\mathbf{r})$ denote the distribution of $\mathbf{r}$ for a given input distribution $f(\mathbf{R})$, and lastly, let $q(\mathbf{r})$ be an arbitrary distribution of $\mathbf{r}$. The mutual information in (24) can be upper-bounded by using duality as [31, Th. 5.1]

$$I(\mathbf{r};\mathbf{R}) = \mathbb{E}_{f(\mathbf{R})}\left[\mathcal{D}\Big(f(\mathbf{r}|\mathbf{R})||f(\mathbf{r})\Big)\right] \tag{28}$$

$$= \mathbb{E}_{f(\mathbf{R})}\left[\mathcal{D}\Big(f(\mathbf{r}|\mathbf{R})||q(\mathbf{r})\Big)\right] - \mathcal{D}\Big(f(\mathbf{r})||q(\mathbf{r})\Big) \tag{29}$$

$$\leq \mathbb{E}_{f(\mathbf{R})}\left[\mathcal{D}\Big(f(\mathbf{r}|\mathbf{R})||q(\mathbf{r})\Big)\right] \tag{30}$$

$$= -\mathbb{E}_{f(\mathbf{r})}\left[\ln\Big(q(\mathbf{r})\Big)\right] - h(\mathbf{r}|\mathbf{R}). \tag{31}$$

Here, $\mathcal{D}(\cdot||\cdot)$ denotes the relative entropy between two probability distributions [29, Eq. 8.46]; (28) follows the definition of the mutual information [29, Eq. 8.49]; in (29) we used Topsøe's identity [32]; (30) follows because of the nonnegativity of relative entropy [29, Thm. 8.6.1]. Finally, (31) follows diretly from the definition of the relative entropy [29, Eq. 8.46].

From (30), we see that any choice of the auxiliary output distribution $q(\mathbf{r})$ results in an upper bound for $I(\mathbf{r};\mathbf{R})$. However, $q(\mathbf{r})$ needs to be selected such that a tight upper bound is obtained. Specifically, we choose the output amplitudes to be independently distributed from $q(r)$, which

is a particular mixture of a half-normal distribution and a Rayleigh distribution

$$q(r) = \frac{\alpha_{\mathrm{U}}(\mu)}{\sqrt{\frac{\sigma_{\mathrm{W}}^2}{(r+\mu)^2} + \sigma_{\Delta}^2}} e^{-\beta_{\mathrm{U}}(\mu) r^2}, \quad r > 0. \tag{32}$$

We will soon motivate the form of $q(r)$. Moreover, in Section VI, we will show that this choice of $q(r)$ results in a tight upper bound on the capacity.

In (32), $\mu \geq 0$ is a constant that will be optimized later to tighten the upper bound. For any μ, the parameters $\alpha_{\mathrm{U}}(\mu)$ and $\beta_{\mathrm{U}}(\mu)$ should be chosen such that certain constraints on $q(r)$ are satisfied. The first constraint is based on the fact that $q(r)$ is a probability distribution function and, hence, must integrate to one

$$\int_0^{\infty} q(r) \mathrm{d}r = 1. \tag{33}$$

The second constraint is due to the input power constraint (8), and can be found from (3)

$$\int_0^{\infty} r^2 q(r) \mathrm{d}r = E_{\mathrm{s}} + 2\sigma_{\mathrm{W}}^2. \tag{34}$$

Although finding closed-form expressions of $\alpha_{\mathrm{U}}(\mu)$ and $\beta_{\mathrm{U}}(\mu)$ is not straightforward, it is possible to determine their values numerically. The numerical method that we used for computing these parameters is presented in Appendix I, and Tab. I contains their computed values for $\mu = 0$ and for various values of σ_{W}^2 and σ_{Δ}^2.

The transition between the half-normal, and the Rayleigh distributions in (32) is based on the values of σ_{W}^2 and σ_{Δ}^2. At high SNR, where the phase noise dominates ($\sigma_{\mathrm{W}}^2 << \sigma_{\Delta}^2$), $q(r)$ is asymptotically half-normal, while it is a Rayleigh distribution for the low SNR values.

This choice of auxiliary output distribution in (32) is motivated as follows: i) The capacity-achieving distribution of the Gaussian channel is a normal distribution [29], and thus the input (and also the output) amplitude follows a Rayleigh distribution, ii) As shown in [14], a tight upper bound for the phase-noise channel at high SNR can be found by using the duality approach, and considering an optimized Gamma distribution as an auxiliary distribution on $|y|^2$. In that case, by following the standard technique for determining the probability density function of a transformed random variable [33, Ch. 5], it is straightforward to show that $r = |y|$ follows a half-normal distribution.

TABLE I: Numerically calculated values of $\alpha_{\mathrm{U}}(\mu=0)$ and $\beta_{\mathrm{U}}(\mu=0)$, when $E_{\mathrm{s}}=1$, and for various σ_{W}^2 and σ_{Δ}^2.

	σ_{W}^2	$\alpha_{\mathrm{U}}(\mu=0)$	$\beta_{\mathrm{U}}(\mu=0)$
$\sigma_{\Delta}^2=10^{-2}$ [rad^2]	5×10^{-2}	0.43	0.88
	5×10^{-3}	0.17	0.73
	5×10^{-4}	0.10	0.59
	5×10^{-5}	0.09	0.53
	5×10^{-6}	0.08	0.51
$\sigma_{\Delta}^2=10^{-3}$ [rad^2]	5×10^{-2}	0.43	0.94
	5×10^{-3}	0.14	0.92
	5×10^{-4}	0.05	0.73
	5×10^{-5}	0.03	0.59
	5×10^{-6}	0.03	0.53

By substituting (32) in (31), we obtain

$$I(\mathbf{r};\mathbf{R}) \leq -n\mathbb{E}\left[\log_2 q(r)\right] - nh(r|R), \tag{35}$$

$$= -n\log_2(\alpha_{\mathrm{U}}(\mu)) + n\frac{\beta_{\mathrm{U}}(\mu)}{\ln(2)}\mathbb{E}\left[r^2\right] + \frac{n}{2}\mathbb{E}\left[\log_2\left(\frac{\sigma_{\mathrm{W}}^2}{(r+\mu)^2}+\sigma_{\Delta}^2\right)\right] - nh(r|R), \tag{36}$$

where in (35), we used that r is independently and identically distributed (iid). From (34), we have $\mathbb{E}\left[r^2\right] = E_{\mathrm{s}} + 2\sigma_{\mathrm{W}}^2$. By substituting (27) and (36) in (24), then (24) in (6), we obtain an upper bound on the capacity as

$$C(\mathsf{SNR}) \leq -\log_2\left(\frac{\alpha_{\mathrm{U}}(\mu)}{2\pi}\right) + \frac{\beta_{\mathrm{U}}(\mu)}{\ln(2)}(E_{\mathrm{s}}+2\sigma_{\mathrm{W}}^2)$$

$$+\sup_{f(R)}\left\{\frac{1}{2}\mathbb{E}\left[\log_2\left(\frac{\sigma_{\mathrm{W}}^2}{(r+\mu)^2}+\sigma_{\Delta}^2\right)\right] - h(r|R) - h\left(N+\Delta|r,R\right)\right\}, \tag{37}$$

$$= -\log_2\left(\frac{\alpha_{\mathrm{U}}(\mu)}{2\pi}\right) + \frac{\beta_{\mathrm{U}}(\mu)}{\ln(2)}(E_{\mathrm{s}}+2\sigma_{\mathrm{W}}^2)$$

$$+\sup_{f(R)}\left\{\mathbb{E}_{f(R)}\left[\frac{1}{2}\mathbb{E}_{f(w_\parallel,w_\perp)}\left[\log_2\left(\frac{\sigma_{\mathrm{W}}^2}{(r+\mu)^2}+\sigma_\Delta^2\right)\right]-h(r|R)-h\left(N+\Delta|r,R\right)\right]\right\}. \tag{38}$$

Finally, the capacity of the phase-noise channel (1) can be bounded as $C(\mathsf{SNR}) \leq C_{\mathrm{U}}(\mathsf{SNR})$, where

$$C_{\mathrm{U}}(\mathsf{SNR}) = \min_{\mu\geq 0}\left\{-\log_2\left(\frac{\alpha_{\mathrm{U}}(\mu)}{2\pi}\right)+\frac{\beta_{\mathrm{U}}(\mu)}{\ln(2)}(E_{\mathrm{s}}+2\sigma_{\mathrm{W}}^2)+\max_{R\geq 0}\mathcal{G}(R)\right\}, \tag{39}$$

and

$$\begin{aligned}\mathcal{G}(R) =& \frac{1}{2}\mathbb{E}_{f(w_\parallel,w_\perp)}\left[\log_2\left(\frac{\sigma_{\mathrm{W}}^2}{\left(\sqrt{(R+w_\parallel)^2+w_\perp^2}+\mu\right)^2}+\sigma_\Delta^2\right)\right]\\ &-h\left(\sqrt{(R+w_\parallel)^2+w_\perp^2}\right)-h\left(\arctan\frac{w_\perp}{R+w_\parallel}+\Delta\middle|r\right).\end{aligned} \tag{40}$$

In (39), the expectation over $f(R)$ is upper-bounded by the maximum value of the expression, which expectation is taken over, for a given R. Finally, the bound can be tightened by minimizing over $\mu \geq 0$.

The upper bound in (37) is further simplified and the final result is presented in the following proposition.

Proposition 2: Capacity of the Wiener phase noise channel (1) is upper-bounded as

$$C(\mathsf{SNR}) \leq C_{\tilde{\mathrm{U}}}(\mathsf{SNR})+o(1), \quad \mathsf{SNR}\to\infty \tag{41}$$

where $o(1)$ denotes a function that vanishes as SNR grows large, and

$$C_{\tilde{\mathrm{U}}}(\mathsf{SNR}) = \frac{\beta_{\mathrm{U}}(\mu=0)}{\ln(2)}(E_{\mathrm{s}}+2\sigma_{\mathrm{W}}^2)-\frac{1}{2}\log_2\sigma_{\mathrm{W}}^2e^2\alpha_{\mathrm{U}}^2(\mu=0). \tag{42}$$

Proof. Please refer to Appendix II. ■

As we shall see from simulation results in Section VI, the provided upper bound is tight for a wide range of SNR values.

V. Capacity Achieving Distributions and the Capacity Lower Bound

One approach to find the capacity lower bound is to restrict the input to have a particular distribution. However, the input distribution must be chosen such that a tight lower bound is

obtained. In this section, we present a method to intelligently choose the distribution of input amplitudes, $f(\mathbf{R})$.

We first reconsider the amplitude and phase channel models in (3) and (4), which at high SNR, reduce to

$$r_k = \sqrt{(R_k + w_{k,\parallel})^2 + w_{k,\perp}^2} \tag{43}$$

$$= \big(R_k + w_{k,\parallel}\big)\sqrt{1 + \frac{w_{k,\perp}^2}{(R_k + w_{k,\parallel})^2}} \tag{44}$$

$$\approx R_k + w_{k,\parallel} \tag{45}$$

$$\theta_k = \Theta_k + \arctan\frac{w_{k,\perp}}{R_k + w_{k,\parallel}} + \phi_k \tag{46}$$

$$\approx \Theta_k + N_k + \phi_k, \tag{47}$$

where $N_k \triangleq w_{k,\perp}/R_k$. In (47), we used that $\arctan(z) \approx z$ for small z. In the following, we study the channel defined in (45) and (47), and derive the capacity-achieving distribution for this simplified channel.

By using the approximate input-output amplitude and phase relations in (45) and (47), and by following the same steps that lead to (24), we obtain

$$\frac{1}{n}I(\mathbf{r},\boldsymbol{\theta};\mathbf{R},\boldsymbol{\Theta}) = \frac{1}{n}I(\mathbf{r};\mathbf{R}) - h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r},\mathbf{R}\right) + \log_2 2\pi \tag{48}$$

$$= \frac{1}{n}\Big(h(\mathbf{r}) - h(\mathbf{r}|\mathbf{R})\Big) - h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r}\right) + \log_2 2\pi \tag{49}$$

$$= \frac{1}{n}\Big(h(\mathbf{r}) - h(\mathbf{w}_{\parallel})\Big) - h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r}\right) + \log_2 2\pi \tag{50}$$

$$= \frac{1}{n}h(\mathbf{r}) - \frac{1}{2}\log_2 2\pi e\sigma_{\mathrm{W}}^2 - h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r}\right) + \log_2 2\pi, \tag{51}$$

where $\mathbf{w}_{\parallel} = \{w_{k,\parallel}\}_{k=1}^n$. Here, (49) follows because $\{a_k\}_{k=1}^n$ is independent of $\mathbf{R}$ given $\mathbf{r}$, and in (51), we used that the entropy of the n-dimensional Gaussian-distributed random variable $\mathbf{w}_{\parallel}$ is given by [29, Thm. 8.4.1]

$$h(\mathbf{w}_{\parallel}) = \frac{n}{2}\log_2 2\pi e\sigma_{\mathrm{W}}^2. \tag{52}$$

Any choice of $f(\mathbf{R})$ results in a lower bound on the capacity of the approximate amplitude-phase channel. Note that the infinite memory of the Wiener phase noise process, ϕ_k, is resolved

by rearranging the noise vector $\mathbf{N}+\boldsymbol{\phi}$ into the difference between the consecutive noise samples. This results in dependency among $\{a_k\}_{k=1}^n$ samples (see (17)-(21)), and motivates to consider an input distribution that introduces a limited order dependency across the amplitudes of the consecutive symbols. More specifically, we consider *block-independent* input amplitudes and confine the optimization in (7) to the set of input distributions of the form

$$\tilde{f}(\mathbf{R}) = \prod_{k=1}^{n/M} f\left(\tilde{\mathbf{R}}^{(k)}\right) = \left(f(\tilde{\mathbf{R}})\right)^{n/M}, \tag{53}$$

where $\tilde{\mathbf{R}}^{(k)}$ are blocks of length-$M > 1$ samples obtained by dividing the vector of input amplitudes [4]

$$\mathbf{R} = \underbrace{\{R_1, \ldots, R_M\}}_{\triangleq \tilde{\mathbf{R}}^{(1)}}, \underbrace{\{R_{M+1}, \ldots, R_{2M}\}}_{\triangleq \tilde{\mathbf{R}}^{(2)}}, \ldots, \underbrace{\{R_{n-M+1}, \ldots, R_n\}}_{\triangleq \tilde{\mathbf{R}}^{(n/M)}}. \tag{54}$$

The second equality in (53) follows as $\tilde{\mathbf{R}}^{(k)}$ are iid from $f(\tilde{\mathbf{R}})$. By substituting (51) in (7), and limiting the input distributions to (53), we obtain

$$C(\mathsf{SNR}) \geq \lim_{n\to\infty} \sup_{f(\tilde{\mathbf{R}})} \left\{ \frac{1}{n} h(\mathbf{r}) - h\left(a_n | \{a_k\}_{k=1}^{n-1}, \mathbf{r}\right) \right\} + \log_2 2\pi - \frac{1}{2}\log_2 2\pi e \sigma_{\mathsf{W}}^2. \tag{55}$$

According to the input-output relation (1), block-independent input symbols result in block-independent output samples denoted by

$$\mathbf{r} = \underbrace{\{r_1, \ldots, r_M\}}_{\triangleq \tilde{\mathbf{r}}^{(1)}}, \underbrace{\{r_{M+1}, \ldots, r_{2M}\}}_{\triangleq \tilde{\mathbf{r}}^{(2)}}, \ldots, \underbrace{\{r_{n-M+1}, \ldots, r_n\}}_{\triangleq \tilde{\mathbf{r}}^{(n/M)}}. \tag{56}$$

Thus, a lower bound on the capacity can be found by using (55) and (56) as

$$\begin{aligned} C(\mathsf{SNR}) &\geq \lim_{n\to\infty} \sup_{f(\tilde{\mathbf{R}})} \left\{ \frac{1}{n} h(\tilde{\mathbf{r}}^{(1)}, \tilde{\mathbf{r}}^{(2)}, \ldots, \tilde{\mathbf{r}}^{(n/M)}) - h\left(a_n | \{a_k\}_{k=1}^{n-1}, \tilde{\mathbf{r}}^{(1)}, \tilde{\mathbf{r}}^{(2)}, \ldots, \tilde{\mathbf{r}}^{(n/M)}\right) \right\} \\ &\quad + \log_2 2\pi - \frac{1}{2}\log_2 2\pi e \sigma_{\mathsf{W}}^2 \qquad (57) \\ &= \lim_{n\to\infty} \sup_{f(\tilde{\mathbf{R}})} \left\{ \frac{1}{n} \left(h(\tilde{\mathbf{r}}^{(1)}) + h(\tilde{\mathbf{r}}^{(2)} | \tilde{\mathbf{r}}^{(1)}) + \ldots \right) - h\left(a_n | \{a_k\}_{k=1}^{n-1}, \tilde{\mathbf{r}}^{(1)}, \tilde{\mathbf{r}}^{(2)}, \ldots, \tilde{\mathbf{r}}^{(n/M)}\right) \right\} \\ &\quad + \log_2 2\pi - \frac{1}{2}\log_2 2\pi e \sigma_{\mathsf{W}}^2 \qquad (58) \end{aligned}$$

[4] Length $M = 1$ is an uninteresting case because a_n depends at least on two consecutive symbols.

$$\geq \lim_{n\to\infty} \sup_{f(\tilde{\mathbf{R}})} \left\{ \frac{1}{M} h(\tilde{\mathbf{r}}^{(n/M)}) - h\left(a_n | \{a_k\}_{k=n-M+2}^{n-1}, \tilde{\mathbf{r}}^{(n/M)}\right) \right\} + \log_2 2\pi - \frac{1}{2}\log_2 2\pi e\sigma_{\mathsf{W}}^2. \tag{59}$$

Here, in (58) the chain rule of entropy is applied to the first term. In (59), we used that $\tilde{\mathbf{r}}^{(k)}$ are iid, and we rewrite the first term. The inequality in (59) holds because removing conditioning from the second term increases entropy. By using the chain rule of entropy on the second term of (59), we obtain

$$C(\mathsf{SNR}) \geq \sup_{f(\tilde{\mathbf{R}})} \left\{ \frac{1}{M} h(\tilde{\mathbf{r}}) + h\left(\{a_k\}_{k=n-M+2}^{n-1} | \tilde{\mathbf{r}}\right) - h\left(a_n, \{a_k\}_{k=n-M+2}^{n-1} | \tilde{\mathbf{r}}\right) \right\} + \log_2 2\pi - \frac{1}{2}\log_2 2\pi e\sigma_{\mathsf{W}}^2, \tag{60}$$

where the superscript of $\tilde{\mathbf{r}}$, and the limit are removed due to stationarity. The second and the third terms on the RHS of (60) correspond to the entropy of two zero-mean Gaussian random vectors with covariance matrices defined as

$$\Sigma_{n-1} \triangleq \mathrm{cov}(\{a_k\}_{k=n-M+2}^{n-1} | \tilde{\mathbf{r}}) \tag{61a}$$

$$\Sigma_{n} \triangleq \mathrm{cov}(\{a_k\}_{k=n-M+2}^{n} | \tilde{\mathbf{r}}). \tag{61b}$$

By using (61) and the definition of the entropy for a Gaussian random variable [29, Thm. 8.4.1], (60) can be rewritten as

$$C(\mathsf{SNR}) \geq \sup_{f(\tilde{\mathbf{R}})} \left\{ \frac{1}{M} h(\tilde{\mathbf{r}}) - \mathbb{E}_{f(\tilde{\mathbf{r}})}\left[\frac{1}{2}\log_2 2\pi e\ g_{\mathbf{r}}(\tilde{\mathbf{r}})\right] \right\} + \log_2 2\pi - \frac{1}{2}\log_2 2\pi e\sigma_{\mathsf{W}}^2, \tag{62}$$

where $g_{\mathbf{r}}(\tilde{\mathbf{r}}) = |\Sigma_n|/|\Sigma_{n-1}|$.

In order to find a tight lower bound on the capacity, we need to find the supremum in (62) by searching over all probability distributions on $\tilde{\mathbf{R}}$ that satisfy the power constraint (8). Unfortunately, this optimization is not mathematically tractable in general. However, the search space reduces at high SNR, since $f(\tilde{\mathbf{r}})$ converges to $f(\tilde{\mathbf{R}})$. Here, determining the optimized $f(\tilde{\mathbf{R}})$ becomes equivalent to finding the optimized $f(\tilde{\mathbf{r}})$. Therefore, we maximize (62) over $f(\tilde{\mathbf{r}})$ by means of functional optimization. The optimized $f(\tilde{\mathbf{r}})$ is used as $f(\tilde{\mathbf{R}})$, which is used to evaluate the lower bound for the capacity of the channel in (1). The optimization steps needed

TABLE II: Numerically calculated values of α_{L} and β_{L}, for $E_{\mathrm{s}}=1$, $\sigma_{\Delta}^2=10^{-3}$ and various σ_{W}^2.

σ_{W}^2	$\alpha_{\mathrm{L}}^{(2)}$	$\beta_{\mathrm{L}}^{(2)}$	$\alpha_{\mathrm{L}}^{(3)}$	$\beta_{\mathrm{L}}^{(3)}$
5×10^{-2}	0.509	0.997	0.12500	0.991
5×10^{-3}	0.051	0.967	0.00400	0.936
5×10^{-4}	0.006	0.825	0.00020	0.756
5×10^{-5}	0.001	0.634	0.00003	0.598
5×10^{-6}	0.001	0.544	0.00002	0.533

are described in Appendix III, and finally, $f(\tilde{\mathbf{R}})$ is found as

$$f(\tilde{\mathbf{R}})=\alpha_{\mathrm{L}}^{(M)}\left(g_{\mathbf{r}}(\tilde{\mathbf{R}})\right)^{-M/2}e^{-\left(\beta_{\mathrm{L}}^{(M)}||\tilde{\mathbf{R}}||^2\right)}, \tag{63}$$

where the parameters $\alpha_{\mathrm{L}}^{(M)}$ and $\beta_{\mathrm{L}}^{(M)}$ are chosen such that the following constraints are satisfied. The first constraint is based on the fact that $f(\tilde{\mathbf{R}})$ is a probability distribution function and, hence, must integrate to one

$$\int_0^{\infty} f(\tilde{\mathbf{R}})\mathrm{d}\tilde{\mathbf{R}}=1. \tag{64}$$

The second constraint is due to the input power constraint (8)

$$\int_0^{\infty} ||\tilde{\mathbf{R}}||^2 f(\tilde{\mathbf{R}})\mathrm{d}\tilde{\mathbf{R}}=ME_{\mathrm{s}}. \tag{65}$$

To numerically compute $\alpha_{\mathrm{L}}^{(M)}$ and $\beta_{\mathrm{L}}^{(M)}$, a method similar to that presented in Appendix I is used. Tab. II contains the calculated values of these parameters in various scenarios. Note that the superscript (M) shows the dependency of these parameters on the dimension of $\tilde{\mathbf{R}}$.

In the numerical result section, we will see that choosing the input distribution as proposed in (63) results in a tight lower bound on the capacity for a wide range of SNR values and phase noise innovation variances.

In the following, we evaluate $f(\tilde{\mathbf{R}})$ for $M=2$ and $M=3$, which will be used in our numerical simulations. For $M=2$, the second term of (59) is equal to $h\left(a_n|\tilde{\mathbf{r}}\right)$. Hence,

$$g_{\mathbf{r}}(\tilde{\mathbf{r}})=|\Sigma_n|=\mathrm{cov}(a_n|\tilde{\mathbf{r}})=\sigma_{\Delta}^2+\frac{\sigma_{\mathrm{W}}^2}{r_n^2}+\frac{\sigma_{\mathrm{W}}^2}{r_{n-1}^2}. \tag{66}$$

By using (63) and (66), we obtain the two-dimensional input distribution as

$$f(R_n, R_{n-1}) = \alpha_{\mathrm{L}}^{(2)} \frac{e^{-\left(\beta_{\mathrm{L}}^{(2)}(R_n^2+R_{n-1}^2)\right)}}{\sigma_\Delta^2 + \frac{\sigma_{\mathrm{W}}^2}{R_n^2} + \frac{\sigma_{\mathrm{W}}^2}{R_{n-1}^2}}. \tag{67}$$

For $M = 3$, the covariance matrices in (61) can be computed as

$$\Sigma_{n-1} = \mathrm{cov}(a_{n-1}|\tilde{\mathbf{r}}) = \sigma_\Delta^2 + \frac{\sigma_{\mathrm{W}}^2}{r_{n-1}^2} + \frac{\sigma_{\mathrm{W}}^2}{r_{n-2}^2}, \tag{68}$$

$$\Sigma_n = \mathrm{cov}(a_n, a_{n-1}|\tilde{\mathbf{r}}) \tag{69}$$

$$= \begin{pmatrix} \sigma_\Delta^2 + \frac{\sigma_{\mathrm{W}}^2}{r_n^2} + \frac{\sigma_{\mathrm{W}}^2}{r_{n-1}^2} & \frac{\sigma_{\mathrm{W}}^2}{r_{n-1}^2} \\ \frac{\sigma_{\mathrm{W}}^2}{r_{n-1}^2} & \sigma_\Delta^2 + \frac{\sigma_{\mathrm{W}}^2}{r_{n-1}^2} + \frac{\sigma_{\mathrm{W}}^2}{r_{n-2}^2} \end{pmatrix}. \tag{70}$$

By using (63) we obtain a three-dimensional input distribution

$$f(R_n, R_{n-1}, R_{n-2}) = \alpha_{\mathrm{L}}^{(3)} \frac{e^{-\left(\beta_{\mathrm{L}}^{(3)}(R_n^2+R_{n-1}^2+R_{n-2}^2)\right)}}{\left(g_{\mathbf{r}}(\tilde{\mathbf{R}})\right)^{-3/2}}, \tag{71}$$

where

$$g_{\mathbf{r}}(\tilde{\mathbf{R}}) = \sigma_\Delta^2 + \frac{\sigma_{\mathrm{W}}^2}{R_n^2} + \frac{\sigma_{\mathrm{W}}^2}{R_{n-1}^2} - \frac{\frac{\sigma_{\mathrm{W}}^4}{R_{n-1}^4}}{\sigma_\Delta^2 + \frac{\sigma_{\mathrm{W}}^2}{R_{n-1}^2} + \frac{\sigma_{\mathrm{W}}^2}{R_{n-2}^2}}. \tag{72}$$

VI. Numerical Results

In this section, we first evaluate the proposed upper bounds (39) and (41). Then, we use the derived input distribution (63) to find lower bounds on the capacity through numerical simulations. We compare the upper bounds with the capacity of the AWGN channel [29, Ch. 9]

$$C_{\mathrm{AWGN}}(\mathsf{SNR}) = \log_2\left(1 + \frac{E_{\mathrm{s}}}{2\sigma_{\mathrm{W}}^2}\right), \tag{73}$$

and the asymptotic high-SNR capacity of the Wiener phase noise channel derived by Lapidoth in [14],

$$C_{\mathrm{Lapidoth}}(\mathsf{SNR}) = \frac{1}{2}\log_2\left(1 + \frac{E_{\mathrm{s}}}{4\sigma_{\mathrm{W}}^2}\right) - \frac{1}{2}\log_2\left(\frac{e\sigma_\Delta^2}{2\pi}\right). \tag{74}$$

Figs. 1 and 2 illustrate the upper bounds (39) and (42) for different SNR values. In Fig. 1, the phase noise innovation variance is $\sigma_\Delta^2 = 10^{-3}$, while it is $\sigma_\Delta^2 = 10^{-2}$ in Fig. 2.

The upper bound $C_{\mathrm{U}}(\mathsf{SNR})$ in (39) is calculated by means of Monte-Carlo simulations. More specifically, we compute $\mathcal{G}(R)$ in (40) for given values of R and μ, by drawing samples

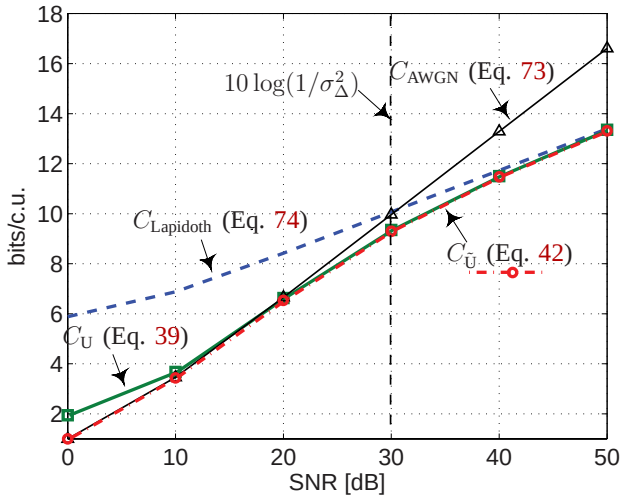

Fig. 1: The proposed capacity upper bound vs. the capacity of the AWGN channel and the asymptotic high-SNR capacity of the phase noise channel from [14]. Here, $\sigma_\Delta^2 = 10^{-3}$ [rad^2].

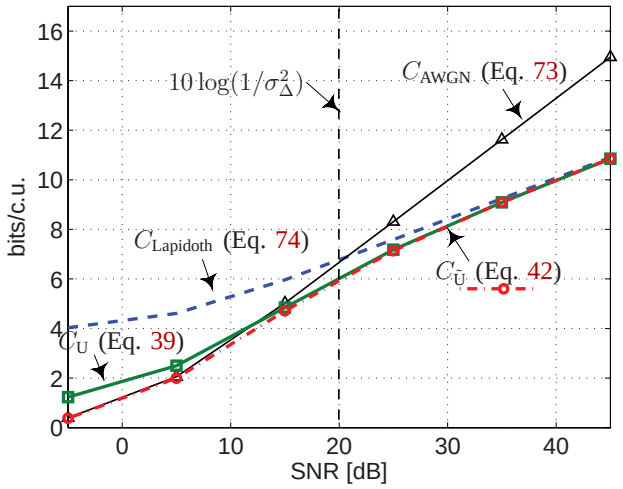

Fig. 2: The proposed capacity upper bound vs. the capacity of the AWGN channel and the asymptotic high-SNR capacity of the phase noise channel from [14]. Here, $\sigma_\Delta^2 = 10^{-2}$ [rad^2].

from $w_\parallel$, $w_\perp$, and Δ. The first term of $\mathcal{G}(R)$ is a double integral, computed numerically. The differential entropy terms of $\mathcal{G}(R)$ are estimated by using the *nearest neighbor estimator* [34].

To evaluate the upper bound in (41), we omit the $o(1)$ term and plot $C_{\tilde{\text{U}}}(\mathsf{SNR})$ from (42). Figs. 1 and 2 show that the this asymptotic upper bound expression $C_{\tilde{\text{U}}}(\mathsf{SNR})$ matches $C_{\text{U}}(\mathsf{SNR})$

for SNR values around 10 dB and above.

It can be seen from Figs. 1 and 2 that the upper bound $C_{\mathrm{U}}(\mathsf{SNR})$ is not tight for SNRs below 10 dB (because it exceeds the AWGN capacity). An alternative upper bound, tighter than $C_{\mathrm{U}}(\mathsf{SNR})$, is $\min\{C_{\mathrm{U}}(\mathsf{SNR}), C_{\mathrm{AWGN}}(\mathsf{SNR})\}$. We observe from the figures that the asymptotic upper bound $C_{\tilde{\mathrm{U}}}(\mathsf{SNR})$ is a very accurate approximation of this alternative bound.

In general, at low SNR, the capacity upper bound approaches the AWGN capacity (73) because the AWGN dominates over the phase noise. At high SNR, where the phase noise dominates, the derived upper bounds follow the high-SNR capacity of the Wiener phase noise channel (74). It can be seen from Figs. 1 and 2 that phase noise starts to dominate for SNR values larger than approximately $10\log(1/\sigma_{\Delta}^2)$ dB. The exact value of this point can also be analytically found by intersecting (73) and (74).

Fig. 3 shows $C_{\tilde{\mathrm{U}}}(\mathsf{SNR})$ for different values of phase noise innovation variance, σ_{Δ}^2. As expected, the bound approaches the capacity of the AWGN channel without phase noise when $\sigma_{\Delta}^2 << \sigma_{\mathrm{W}}^2$.

In order to numerically find a lower bound on the capacity of the channel (1), the sum-product algorithm proposed in [35] for calculation of the information rate of channels with memory is used. We specifically use the particle-based implementation of this method, which is introduced in [36]. First, we use the rejection sampling approach [37] to draw samples from (63) for the input amplitudes. For the phase of the input samples, we use Lemma 1, and draw independent samples from $\mathcal{U}(0, 2\pi)$. The generated input samples are transmitted over the original channel (1), and the achievable information rate is computed as explained in [36].

Fig. 4 compares the simulated lower bounds for $M = 2$ and $M = 3$ against the calculated upper bound. The particle-based method of [36] with 10^7 particles over 10^3 channel uses is employed. It can be seen that the computed lower bounds are close to the upper bound for a wide range of SNR values. In particular, the input with a higher order of dependency of amplitudes $(M = 3)$ results in a tighter lower bound at low SNR.

VII. Conclusion

In this paper, we presented methods to develop tight upper and lower bounds on the capacity of the Wiener phase noise channel. A capacity upper bound, tighter than that of available in the literature, is derived. We also derived analytical expressions for a family of input distributions,

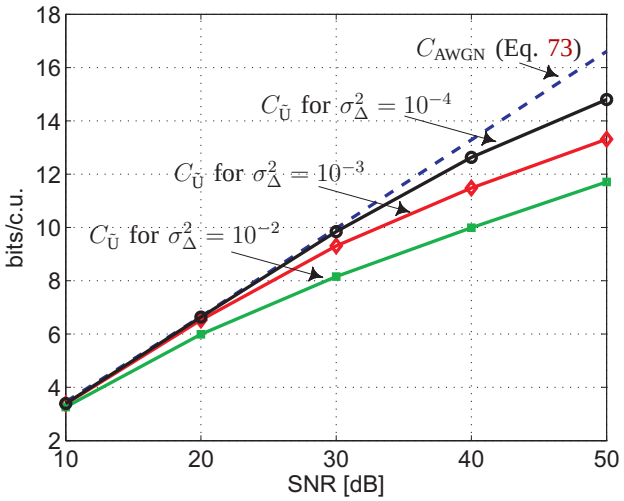

Fig. 3: The proposed capacity upper bound for various σ_Δ^2, compared against the AWGN capacity.

which result in tight lower bounds on the capacity. The proposed input distributions are circularly symmetric and non-Gaussian. Moreover, the input amplitudes are correlated over time. The proposed upper and lower bounds tightly enclose the channel capacity for a wide range of SNR values. The proposed bounds reach the AWGN capacity at low SNR. In the limiting regime of high SNR, only the amplitude of the transmitted signal can be perfectly recovered, whereas the phase is lost. Therefore in that regime, by increasing the SNR gains in capacity can only be achieved through the amplitude channel.

APPENDIX I
NUMERICAL CALCULATION OF $\alpha_{\text{U}}(\mu)$ AND $\beta_{\text{U}}(\mu)$

In this section, we present the numerical method that is used for the calculation of $\alpha_{\text{U}}(\mu)$ and $\beta_{\text{U}}(\mu)$. From (33) and (34), we obtain

$$\frac{\int_0^\infty r^2 q(r)\mathrm{d}r}{\int_0^\infty q(r)\mathrm{d}r} = E_{\text{s}} + 2\sigma_{\text{W}}^2, \tag{75}$$

where the left hand side of (75) is independent of $\alpha_{\text{U}}(\mu)$ based on (32). To find $\beta_{\text{U}}(\mu)$ that satisfies (75), we use the bisection method. In each iteration of the algorithm, the integrals involved in (75) are numerically calculated. After finding $\beta_{\text{U}}(\mu)$, (33) is used to compute $\alpha_{\text{U}}(\mu)$.

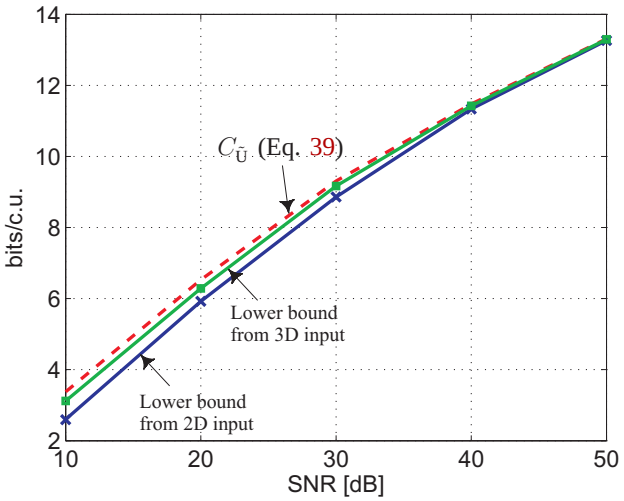

Fig. 4: Lower bounds computed by using the proposed 2D and 3D input distributions compared with the proposed capacity upper bound. In this figure $\sigma_{\Delta}^2 = 10^{-3}$.

APPENDIX II
PROOF OF PROPOSITION 2

To further simplify the upper bound in (39), we use the *escape-to-infinity* property of the capacity-achieving input distribution [31, Def. 4.11]. We impose the additional constraint $R \geq R_0$ on the input distribution, where $R_0 > 0$, and denote the capacity of the channel (1) as $C^{R_0}(\mathsf{SNR})$. By following the same procedure as in [31, Th. 4.12] and [38, Th. 8], we can show that

$$C(\mathsf{SNR}) = C^{R_0}(\mathsf{SNR}) + o(1), \quad \mathsf{SNR} \to \infty, \tag{76}$$

where $o(1)$ denotes a function that vanishes as SNR grows large. This means that the high-SNR behavior of $C(\mathsf{SNR})$ does not change if the input distribution is constrained to lie outside a sphere of an arbitrary radius. Considering this new constraint and repeating the steps leading to (39), we obtain

$$C^{R_0}(\mathsf{SNR}) \leq \min_{\mu \geq 0} \left\{ -\log_2\left(\frac{\alpha_{\mathrm{U}}(\mu)}{2\pi}\right) + \frac{\beta_{\mathrm{U}}(\mu)}{\ln(2)}(E_{\mathrm{s}} + 2\sigma_{\mathrm{W}}^2) + \max_{R \geq R_0} \mathcal{G}(R) \right\}, \tag{77}$$

where $\mathcal{G}(R)$ is defined in (40). By choosing R_0 to be arbitrary large, we can evaluate $\mathcal{G}(R)$ when $R \to \infty$. The first term is found as

$$\lim_{R \to \infty} \frac{1}{2} \mathbb{E}_{f(w_{\parallel}, w_{\perp})} \left[\log_2 \left(\frac{\sigma_{\mathrm{W}}^2}{\left(\sqrt{(R + w_{\parallel})^2 + w_{\perp}^2} + \mu\right)^2} + \sigma_{\Delta}^2 \right) \right] = \frac{1}{2} \log_2\left(\sigma_{\Delta}^2\right). \tag{78}$$

The equality in (78) follows from the dominated convergence theorem that permits interchange of the limit and the expectation operators.[5]

For the second term of $\mathcal{G}(R)$ we obtain

$$\lim_{R\to\infty} h\left(\sqrt{(R+w_{\|})^2+w_{\perp}^2}\right) = \lim_{R\to\infty} h\left(R+w_{\|}+O\left(\frac{1}{R}\right)\right) \tag{79}$$

$$= \lim_{R\to\infty} h\left(R+w_{\|}\right) \tag{80}$$

$$= \frac{1}{2}\log_2 2\pi e\sigma_{\mathrm{W}}^2. \tag{81}$$

In (79), we have written the power series of the function inside the differential entropy about $R=\infty$. The $O(1/R)$ term represents omitted terms of order $1/R$. In (80), we used [31, Lemma. 6.9]

$$\lim_{\epsilon\to 0} h(a+\epsilon b) = h(a). \tag{82}$$

Finally, in (81), we used that translation does not change differential entropy. Then, we used the entropy of a Gaussian distributed random variable [29, p. 244].

Similarly, for the third term of $\mathcal{G}(R)$, we obtain

$$\lim_{R\to\infty} h\left(\arctan\frac{w_{\perp}}{R+w_{\|}}+\Delta\middle| r\right) = h\left(\Delta\right) \tag{83}$$

$$= \frac{1}{2}\log_2 2\pi e\sigma_{\Delta}^2. \tag{84}$$

Here in (84), we again used (82), and the fact that

$$\lim_{R\to\infty} \arctan\frac{w_{\perp}}{R+w_{\|}} = 0. \tag{85}$$

Substituting (78), (81), and (84) in (77), we obtain

$$C^{R_0}(\mathsf{SNR}) \leq \min_{\mu\geq 0}\left\{\frac{\beta_{\mathrm{U}}(\mu)}{\ln(2)}(E_{\mathrm{s}}+2\sigma_{\mathrm{W}}^2) - \frac{1}{2}\sigma_{\mathrm{W}}^2 e^2 \log_2 \alpha_{\mathrm{U}}^2(\mu)\right\}. \tag{86}$$

From (32) we observe that $q(r)$ becomes asymptotically independent of μ when $R\to\infty$, and thus become $\alpha_{\mathrm{U}}(\mu)$ and $\beta_{\mathrm{U}}(\mu)$. Hence in this case, $\alpha_{\mathrm{U}}(\mu)$ and $\beta_{\mathrm{U}}(\mu)$ can be found for any arbitrary μ, and the minimization in (86) can be omitted. Without loss of generality, we set $\mu=0$, and substitute (86) in (76)

$$C(\mathsf{SNR}) \leq \frac{\beta_{\mathrm{U}}(\mu=0)}{\ln(2)}(E_{\mathrm{s}}+2\sigma_{\mathrm{W}}^2) - \frac{1}{2}\log_2 \sigma_{\mathrm{W}}^2 e^2 \alpha_{\mathrm{U}}^2(\mu=0) + o(1). \tag{87}$$

This concluded the proof of Proposition 2.

[5]It is straightforward to show that the dominated convergence theorem holds by choosing an integrable dominated (majorizer) function such as $\log_2\left(\sigma_{\mathrm{W}}^2/w_{\perp}^2+\sigma_{\Delta}^2\right)$ over the function inside the expectation.

APPENDIX III
OPTIMIZED INPUT DISTRIBUTION

A. Background on Functional Optimization

Definition of the capacity stated in (6) is a functional optimization problem [39]. A functional is a mapping from a function to a real number, e.g. an integral. A background on *calculus of variations* [39] is presented here that later will be used in our analysis.

Consider the functional $F(u)$, defined as

$$F(u) = \int_{\Omega} K(\mathbf{x}, u(\mathbf{x}))\, \mathrm{d}\mathbf{x}, \tag{88}$$

where $u(\mathbf{x})$ is a real valued function of a real vector argument $\mathbf{x}$,

$$u : \Omega \subset \mathbb{R}^N \to \mathbb{R}. \tag{89}$$

A necessary condition for u to be a stationary point of $F(u)$ under the m constraints

$$\int_{\Omega} L_i(\mathbf{x}, u(\mathbf{x}))\, \mathrm{d}\mathbf{x} = 0,\ i = 1, 2, ..., m\,, \tag{90}$$

is that the following simplified Euler-Lagrange equation is satisfied

$$\frac{\partial K}{\partial u} + \sum_{i=1}^{m} \lambda_i \frac{\partial L_i}{\partial u} = 0. \tag{91}$$

The Lagrange multipliers λ_i should be chosen to fulfill the constraints. Note that, K and L_i, $i = 1, 2, ..., m$, are real valued functions with continuous first partial derivatives.

B. Functional Optimization of the Lower Bound

We perform the supremum by employing the functional optimization method reviewed in Appendix III-A. Based on (62), the functional that must be maximized is written as

$$F(f(\tilde{\mathbf{r}})) = \int_0^{\infty} \Big[-\frac{1}{M} f(\tilde{\mathbf{r}}) \log(f(\tilde{\mathbf{r}})) - f(\tilde{\mathbf{r}}) \frac{1}{2} \log_2 2\pi e\ g_{\mathbf{r}}(\tilde{\mathbf{r}}) + \log_2 2\pi - \frac{1}{2} \log_2 2\pi e \sigma_{\mathrm{W}}^2 \Big] \mathrm{d}\tilde{\mathbf{r}}. \tag{92}$$

Using (92), the function corresponding to K in (88) is identified as

$$K(\tilde{\mathbf{r}}, f(\tilde{\mathbf{r}})) = -\frac{1}{M} f(\tilde{\mathbf{r}}) \log_2(f(\tilde{\mathbf{r}})(g_{\mathbf{r}}(\tilde{\mathbf{r}}))^{M/2}) - \frac{1}{2} \log_2 \sigma_{\mathrm{W}}^2 e^2. \tag{93}$$

There are two constraints on $f(\tilde{\mathbf{r}})$ that must be satisfied. First, $f(\tilde{\mathbf{r}})$ has to integrate to one. The second constraint can be formulated based on the average power constraint of the input distribution. Consequently, for L_i in (90), we obtain

$$L_1(\tilde{\mathbf{r}}, f(\tilde{\mathbf{r}})) = f(\tilde{\mathbf{r}}) - 1 \tag{94a}$$

$$L_2(\tilde{\mathbf{r}}, f(\tilde{\mathbf{r}})) = ||\tilde{\mathbf{r}}||^2 f(\tilde{\mathbf{r}}) - M(E_{\mathrm{s}} + \sigma_{\mathrm{w}}^2). \tag{94b}$$

Finally, by substituting (93) and (94) into the Euler-Lagrange equation (91), we obtain the output distribution that maximizes (62) as

$$f(\tilde{\mathbf{r}}) = \alpha_{\mathrm{L}}^{(M)} (g_{\mathbf{r}}(\tilde{\mathbf{r}}))^{M/2} e^{-(\beta_{\mathrm{L}}^{(M)} ||\tilde{\mathbf{r}}||^2)}. \tag{95}$$

As mentioned before, the optimized $f(\tilde{\mathbf{r}})$ is used as $f(\tilde{\mathbf{R}})$ to evaluate the lower bound for the capacity of the channel in (1).

ACKNOWLEDGMENT

Discussions with Ashkan Panahi are gratefully acknowledged.

REFERENCES

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

[2] 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, pp. 1748–1757, Sep. 2005.

[3] M.R. Khanzadi, H. Mehrpouyan, E. Alpman, T. Svensson, D. Kuylenstierna, and T. Eriksson, "On models, bounds, and estimation algorithms for time-varying phase noise," *in Proc. Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, pp. 1 –8, Dec. 2011.

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

[5] M.R. Khanzadi, R. Krishnan, and T. Eriksson, "Effect of synchronizing coordinated base stations on phase noise estimation," *in Proc. IEEE Acoust., Speech, Signal Process. (ICASSP)*, pp. 4938–4942, May. 2013.

[6] M.R. Khanzadi, D. Kuylenstierna, A. Panahi, T. Eriksson, and H. Zirath, "Calculation of the performance of communication systems from measured oscillator phase noise," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 5, pp. 1553–1565, May 2014.

[7] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[8] 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," *in Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, pp. 1 –6, Apr. 2012.

[9] R. Krishnan, A. Graell i Amat, T. Eriksson, and G. Colavolpe, "Constellation optimization in the presence of strong phase noise," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5056–5066, Dec. 2013.

[10] P. Smulders, "Exploiting the 60 GHz band for local wireless multimedia access: prospects and future directions," *IEEE Commun. Mag.*, vol. 40, no. 1, pp. 140–147, Aug. 2002.

[11] Y. Li, H. Jacobsson, M. Bao, and T. Lewin, "High-frequency SiGe MMICs-an industrial perspective," *in Proc. GigaHertz 2003 Symp.*, Linköping, Sweden, Nov. 2003.

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

[13] H. Mehrpouyan, M.R. Khanzadi, M. Matthaiou, A. Sayeed, R. Schober, and Y. Hua, "Improving bandwidth efficiency in E-band communication systems," *IEEE Commun. Mag.*, vol. 52, no. 3, pp. 121–128, Mar. 2014.

[14] A. Lapidoth, "On phase noise channels at high SNR," *in Proc. Inf. Theory Workshop (ITW)*, pp. 1–4, Oct. 2002.

[15] G. Colavolpe and R. Raheli, "The capacity of the noncoherent channel," *Europ. Trans. Telecommun.*, vol. 12, no. 4, pp. 289–296, Jul./Aug. 2001.

[16] M. Peleg and S. Shamai (Shitz), "On the capacity of the blockwise incoherent MPSK channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 603–609, May 1998.

[17] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.

[18] G. Durisi, "On the capacity of the block-memoryless phase-noise channel," *IEEE Commun. Lett.*, vol. 16, no. 8, pp. 1157–1160, Aug. 2012.

[19] P. Hou, B. Belzer, and T. Fischer, "Shaping gain of the partially coherent additive white Gaussian noise channel," *IEEE Commun. Lett.*, vol. 6, no. 5, pp. 175–177, May 2002.

[20] H. Ghozlan and G. Kramer, "On Wiener phase noise channels at high signal-to-noise ratio," *in Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2279–2283, Jul. 2013.

[21] ——, "Phase modulation for discrete-time Wiener phase noise channels with oversampling at high SNR," *in Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1554–1557, Jun. 2014.

[22] L. Barletta and G. Kramer, "On continuous-time white phase noise channels," *in Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2426–2429, Jun. 2014.

[23] G. Durisi, A. Tarable, C. Camarda, and G. Montorsi, "On the capacity of MIMO Wiener phase-noise channels," *in Proc. Inf. Theory Applicat. Workshop (ITA)*, Feb. 2013.

[24] 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.*, Sep. 2013, submitted for publication.

[25] M.R. Khanzadi, G. Durisi, and T. Eriksson, "Capacity of SIMO and MISO phase-noise channels with common/separate oscillators," *IEEE Trans. Commun.*, Feb. 2015. [Online]. Available: http://arxiv.org/abs/1409.0561

[26] 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.

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

[28] M.R. Khanzadi, R. Krishnan, D. Kuylenstierna, and T. Eriksson, "Oscillator phase noise and small-scale channel fading in higher frequency bands," *To appear in Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2014.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., New York, NY, U.S.A., 2006.

[30] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2716–2755, Jun. 2009.

[31] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[32] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Scientiarum Math. Hung.*, vol. 2, pp. 291–292, 1967.

[33] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. New York: McGraw-Hill, 2002.

[34] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. Van der Meulen, "Nonparametric entropy estimation: An overview," *Int. J. Math. Statist. Sci.*, vol. 6, no. 1, pp. 17–39, 1997.

[35] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[36] J. Dauwels and H.-A. Loeliger, "Computation of information rates by particle methods," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 406–409, 2008.

[37] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, 2nd ed. New York, NY: Springer-Verlag, 2004.

[38] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.

[39] F. Byron and R. Fuller, *Mathematics of Classical and Quantum Physics*. Dover Publications, 1992.