(Article begins on next page)

On the capacity of MIMO Wiener phase-noise channels

Giuseppe Durisi[1], Alberto Tarable[2], Christian Camarda[3], and Guido Montorsi[3]

[1]Chalmers University of Technology, 41296 Gothenburg, Sweden
[2]IEIIT-CNR, 10129 Turin, Italy
[3]Politecnico di Torino, 10129 Turin, Italy

***Abstract*—The capacity of multiple-antenna systems affected by Wiener phase noise is investigated. We present a non-asymptotic capacity upper bound that is shown to be tight in the large-SNR regime. The capacity upper bound is compared with a lower bound obtained by evaluating numerically the information rates achievable with QAM constellations. For a Wiener phase-noise process with standard deviation of the phase increments equal to 6°, our results suggest that QAM constellations incur a penalty of more than $3\,\mathrm{dB}$ for medium/high SNR values.**

I. Introduction

Phase noise caused by both phase and frequency instabilities in the radio-frequency (RF) oscillators used in wireless transceivers is one of the major impairments in certain communication systems. One example is high-throughput microwave links used to provide backhaul connectivity in wireless cellular networks [1]. These links typically employ high-order constellations (512 QAM is used in commercial products), which make them extremely sensitive to phase noise. Another example are communication systems employing low-cost low-quality RF oscillators, such as in DVB-S2 transceivers (see [2] and references therein) and in the large-MIMO transceivers currently under theoretical investigation [3].

A fundamental way to characterize the impact of phase noise on the throughput of these systems is to study their Shannon capacity. Unfortunately, the capacity of the phase-noise channel is not known in closed-form, even for simple channel models. Lapidoth [4] obtained a large-SNR characterization of the capacity of the general class of stationary phase-noise channels (the widely used Wiener model [2] belongs to this class). Specifically, he showed that the capacity of the phase-noise channel is asymptotically equal to half the capacity of an AWGN channel with the same SNR plus a correction term that accounts for the memory in the phase-noise process. The result in [4] has been recently extended to the *waveform* phase-noise channel in [5]. The capacity of the block-memoryless phase-noise channel (a non-stationary channel) has been characterized in [6] in the large-SNR regime.

Moving away from asymptotic results, Katz and Shamai [7] provided tight upper and lower bounds on the capacity of memoryless phase noise channels. These bounds have been recently extended to the block-memoryless phase-noise case in [8]. For the Wiener phase-noise model, an upper bound on the rates achievable with PSK constellations has been recently proposed in [9]. Capacity lower bounds obtained by numerically computing the information rates achievable with various families of finite-cardinality independent and identically distributed (i.i.d.) input processes (e.g., QAM, PSK, and APSK constellations) have been reported in [9], [2], [10]. The numerical evaluation of these bounds is based on the algorithm for the computation of the information rates for finite-state channels proposed in [11].

The impact of phase noise on multiple-antenna systems has been recently discussed in [1] where it is shown that different RF circuitries configurations (e.g., independent oscillators at each antenna as opposed to a single oscillator driving all antennas) yield different capacity behavior at high SNR.

Contributions: We study the capacity of multiple-antenna systems affected by phase noise. Specifically, we consider the scenario where a single oscillator drives all RF circuitries at each transceiver. We present a non-asymptotic capacity upper bound for the case of Wiener phase noise. When particularized to constant modulus constellations and to single-antenna systems, our bound recovers the upper bound obtained in [9]. By exhibiting a matching lower bound, we show that our upper bound is tight in the large-SNR regime. Focusing on single-antenna systems, we finally compare our upper bound with lower bounds obtained by evaluating numerically the information rates achievable with QAM constellations. For the case of a Wiener phase-noise process with standard deviation of the phase increments equal to 6°, our results imply that QAM constellations incur a penalty of more than $3\,\mathrm{dB}$ for medium/high SNR values.

II. System Model

We consider an $M \times M$ MIMO phase noise channel with memory, described by the following input-output relation

$$\mathbf{y}_k = e^{j\theta_k}\mathbf{H}\mathbf{x}_k + \mathbf{w}_k, \quad k = 1, 2, \dots \tag{1}$$

Here, $\mathbf{x}_k$ denotes the M-dimensional input vector at discrete time k, $\mathbf{H}$ is the MIMO channel matrix, which we assume deterministic, full-rank, and known to the transmitter and the receiver, $\{\theta_k\}$ is the phase-noise process, and $\mathbf{w}_k$ is the additive Gaussian noise, which is circularly symmetric with zero mean and covariance matrix $\mathbf{I}_M$, i.e., $\mathbf{w}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. This model is accurate for MIMO systems where the distance between the antennas at the transceivers is sufficiently small for the RF circuitries at each antenna to be driven by the same oscillator [1].

A common model for the phase-noise process $\{\theta_k\}$ is the Wiener model [2], according to which[1]

$$\theta_{k+1} = \theta_k + \Delta_k \tag{2}$$

where the process $\{\Delta_k\}$ is made of i.i.d. zero-mean Gaussian random variables with variance σ_Δ^2, i.e., $\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2)$. The i.i.d. assumption on $\{\Delta_k\}$ implies that $\{\theta_k\}$ is a Markov process. Specifically,

$$f_{\theta_k \,|\, \theta_{k-1},\dots,\theta_0} = f_{\theta_k \,|\, \theta_{k-1}} = f_\Delta$$

where

$$f_\Delta(\delta) \triangleq \sum_{l=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\Delta^2}} \exp\left(-\frac{(\delta - 2\pi l)^2}{2\sigma_\Delta^2}\right), \quad \delta \in [0, 2\pi]. \tag{3}$$

In words, f_Δ is the probability density function (pdf) of the innovation Δ_k modulo 2π.

Under the additional assumption that θ_0 is uniformly distributed in the interval $[0, 2\pi]$, i.e., $\theta_0 \sim \mathcal{U}[0, 2\pi]$, the process $\{\theta_k\}$ is stationary. Let $\Delta \sim f_\Delta$ (defined in (3)). The differential entropy rate of a stationary Wiener process is

$$h(\{\theta_k\}) = h(\Delta) \leq \frac{1}{2}\log(2\pi e \sigma_\Delta^2).$$

The upper bound, which holds because the Gaussian distribution maximizes differential entropy under a variance constraint [12, Thm. 8.65], turns out to be tight whenever $\sigma_\Delta \lesssim 50°$.

III. Capacity

We are interested in computing the capacity of the MIMO phase-noise channel (1), which is defined as

$$C(\rho) = \lim_{n\to\infty} \frac{1}{n} \sup I(\mathbf{x}^n; \mathbf{y}^n). \tag{4}$$

Here, the supremum is over all probability distributions on $\mathbf{x}^n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ that satisfy the average-power constraint

$$\sum_{k=1}^{n} \mathbb{E}\left[\|\mathbf{x}_k\|^2\right] \leq n\rho. \tag{5}$$

Since the additive noise has unit variance, the parameter $\rho \geq 0$ can be thought of as the SNR. The capacity $C(\rho)$ is not known in closed form. In Section IV we shall present a capacity upper bound that will turn out to be tight in the large-SNR regime.

Before presenting our upper bound, two observations are in order.

i) As $\mathbf{H}$ is known to transmitter and receiver, $C(\rho)$ depends on $\mathbf{H}$ only through its singular values. For simplicity, in the remainder of the paper we shall focus on the special case when all eigenvalues of $\mathbf{H}$ are equal to one. In this case, we can (and will) assume without loss of generality $\mathbf{H} = \mathbf{I}_M$.

ii) The following proposition establishes that the capacity-achieving input process $\{\mathbf{x}_k\}$ can be assumed isotropically distributed, a property that will be useful in our analysis.

[1]See [5] for a discussion on the limitations of this model.

Proposition 1: The input process $\{\mathbf{x}_k\}$ that achieves the capacity of the channel (1), with $\mathbf{H} = \mathbf{I}_M$, can be assumed isotropically distributed. Specifically, if $\{\mathbf{x}_k\}$ achieves $C(\rho)$ in (4) then $\{\mathbf{U}_k\mathbf{x}_k\}$, where the matrix-valued random process $\{\mathbf{U}_k\}$ is i.i.d. and each $\mathbf{U}_k$ is uniformly distributed on the set of $M \times M$ unitary matrices, achieves $C(\rho)$ as well.

Proof: The proof, which exploits that $\mathbf{U}_k\mathbf{w}_k \sim \mathbf{w}_k$, follows the same steps as the proof of [13, Prop. 7]. ■

IV. A Capacity Upper Bound

We next present an upper bound on $C(\rho)$, which is constructed by extending to the MIMO case the method used in [4] to derive an asymptotic bound on the capacity of stationary single-antenna phase-noise channels. We also use the approach proposed in [8] to make the bound non-asymptotic.

Theorem 2: The capacity of the channel (1) can be upper-bounded as $C(\rho) \leq U(\rho)$, where

$$U(\rho) \triangleq \min_{\alpha>0}\min_{\lambda\geq 0}\left\{\alpha \log\frac{\rho+M}{\alpha} + d_{\lambda,\alpha} + \log(2\pi) + \max_{\xi\geq 0} g_{\lambda,\alpha}(\xi,\rho)\right\}. \tag{6}$$

Here,

$$\begin{aligned} g_{\lambda,\alpha}(\xi,\rho) \triangleq{}& (M-\alpha)\,\mathbb{E}\left[\log\left(|\xi + z_1|^2 + \sum_{j=2}^{M}|z_j|^2\right)\right] \\ &- h(|\xi+z|^2) + (\alpha-\lambda)\frac{\xi^2+M}{\rho+M} \\ &- h(\underline{/\xi+z} + \Delta \,\big|\, |\xi+z|) \end{aligned} \tag{7}$$

where $\underline{/x}$ denotes the phase of $x \in \mathbb{C}$ and $z, z_1, \dots, z_M$ are i.i.d. $\mathcal{CN}(0,1)$-distributed random variables. Furthermore,

$$d_{\lambda,\alpha} \triangleq \log\frac{\Gamma(\alpha)}{\Gamma(M)} + \lambda - M + 1. \tag{8}$$

Proof: Because of Proposition 1, we can restrict ourselves to input processes that are isotropically distributed. Specifically, we will consider $\{\mathbf{x}_k\}$ of the form $\{\mathbf{x}_k = s_k\mathbf{v}_k\}$, where $s_k = \|\mathbf{x}_k\|$ and $\mathbf{v}_k = \mathbf{x}_k/s_k$, with $\mathbf{v}_k$ uniformly distributed on the unit sphere in $\mathbb{C}^M$ and independent of s_k. We start by using chain rule as follows

$$I(\mathbf{x}_1^n; \mathbf{y}_1^n) = \sum_{k=1}^{n} I(\mathbf{x}_1^n; \mathbf{y}_k \,|\, \mathbf{y}_1^{k-1}). \tag{9}$$

By proceeding as in [4], we next upper-bound each term on the right-hand side (RHS) of (9) as

$$\begin{aligned} I(\mathbf{x}_1^n; \mathbf{y}_k \,|\, \mathbf{y}^{k-1}) &= h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}) - h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}, \mathbf{x}^n) \\ &\overset{(a)}{\leq} h(\mathbf{y}_k) - h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}, \mathbf{x}^n) \\ &= h(\mathbf{y}_k) - h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k) \\ &\overset{(b)}{\leq} h(\mathbf{y}_k) - h(\mathbf{y}_k \,|\, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}, \mathbf{x}_k, \theta^{k-1}) \\ &\overset{(c)}{=} h(\mathbf{y}_k) - h(\mathbf{y}_k \,|\, \mathbf{x}_k, \theta^{k-1}) \\ &= h(\mathbf{y}_k) - h(\mathbf{y}_k \,|\, \mathbf{x}_k) + h(\mathbf{y}_k \,|\, \mathbf{x}_k) \end{aligned}$$

$$
\begin{aligned}
&- h(\mathbf{y}_k \,|\, \mathbf{x}_k, \theta^{k-1}) \\
&= I(\mathbf{x}_k; \mathbf{y}_k) + I(\mathbf{y}_k; \theta^{k-1} \,|\, \mathbf{x}_k) \\
&\overset{(d)}{=} I(\mathbf{x}_k; \mathbf{y}_k) + I(\mathbf{y}_k; \theta_{k-1} \,|\, \mathbf{x}_k).
\end{aligned} \tag{10}
$$

Here, in (a) and (b) we used that conditioning reduces entropy; (c) follows because $\mathbf{y}_k$ and the pair $(\mathbf{y}^{k-1}, \mathbf{x}^{k-1})$ are conditionally independent given $(\theta^{k-1}, \mathbf{x}_k)$; finally, (d) holds because $\{\theta_k\}$ is a first-order Markov process. Let $z_k \sim \mathcal{CN}(0,1)$. The second term on the RHS of (10) can be evaluated as follows:

$$
\begin{aligned}
I(\mathbf{y}_k; \theta_{k-1} \,|\, \mathbf{x}_k) &\overset{(a)}{=} I(e^{j\theta_k} s_k + z_k; \theta_{k-1} \,|\, s_k) \\
&\overset{(b)}{=} I\big(e^{j\theta_k}(s_k + z_k); \theta_{k-1} \,|\, s_k\big) \\
&= I\big(|s_k + z_k|\,, \underline{/s_k + z_k} + \theta_k; \theta_{k-1} \,|\, s_k\big) \\
&\overset{(c)}{=} I\big(\underline{/s_k + z_k} + \theta_k; \theta_{k-1} \,\big|\, |s_k + z_k|\,, s_k\big) \\
&= h\big(\underline{/s_k + z_k} + \theta_k \,\big|\, |s_k + z_k|\,, s_k\big) \\
&\quad - h\big(\underline{/s_k + z_k} + \theta_k \,\big|\, |s_k + z_k|\,, \theta_{k-1}, s_k\big) \\
&\overset{(d)}{=} \log(2\pi) \\
&\quad - h\big(\underline{/s_k + z_k} + \Delta \,\big|\, |s_k + z_k|\,, s_k\big).
\end{aligned} \tag{11}
$$

Here, (a) follows because $\mathbf{v}_k^{\mathsf{H}} \mathbf{y}_k \sim e^{j\theta_k} s_k + z_k$ is a sufficient statistics for θ_{k-1}; (b) follows because z_k is circularly symmetric; (c) holds because $|s_k + z_k|$ and θ_{k-1} are independent; finally, (d) holds because $\theta_k \sim \mathcal{U}[0, 2\pi]$ and because of (2). Substituting (11) into (10), then (10) into (9), and using that $\{\theta_k\}$ is a stationary process, we get

$$
\begin{aligned}
C(\rho) \le \sup_{\mathcal{Q}_s} \Big\{ & I(s, \mathbf{v}; \mathbf{y}) \\
& + \log(2\pi) - h\big(\underline{/s + z} + \Delta \,\big|\, |s + z|\,, s\big) \Big\}
\end{aligned} \tag{12}
$$

where

$$
\mathbf{y} = e^{j\theta} s \mathbf{v} + \mathbf{w} \tag{13}
$$

with $\theta \sim \mathcal{U}[0, 2\pi]$, $\mathbf{v}$ uniformly distributed on the unit sphere in $\mathbb{C}^M$, $z \sim \mathcal{CN}(0,1)$, and Δ distributed as in (3); the supremum in (12) is over all distributions $\mathcal{Q}_s$ on $s \ge 0$ that satisfy $\mathbb{E}\big[s^2\big] \le \rho$. We further upper-bound the first term on the RHS of (12) (which corresponds to the mutual information achievable on a memoryless channel with uniform phase noise) by using duality [14, Thm. 5.1] and obtain that for every $\mathcal{Q}_s$ and for every $\alpha > 0$ and $\lambda > 0$ (see Appendix)

$$
\begin{aligned}
I(s, \mathbf{v}; \mathbf{y}) \le{} & \alpha \log \frac{\rho + M}{\alpha} + d_{\lambda,\alpha} \\
& + (M - \alpha)\, \mathbb{E}\left[\log\left(|s + z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right] \\
& - h(|s + z|^2 \;|\; s) + (\alpha - \lambda) \frac{\mathbb{E}\big[s^2\big] + M}{\rho + M}.
\end{aligned} \tag{14}
$$

Here, $d_{\lambda,\alpha}$ is the constant defined in (8) and $z, z_1, \dots, z_M$ are i.i.d. $\mathcal{CN}(0,1)$-distributed random variables. Substituting (14) into (12), we obtain

$$
\begin{aligned}
C(\rho) \le{} & \alpha \log \frac{\rho + M}{\alpha} + d_{\lambda,\alpha} + \log(2\pi) \\
& + \sup_{\mathcal{Q}_s} \bigg\{ (M - \alpha)\, \mathbb{E}\left[\log\left(|s + z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right] \\
& \quad - h(|s + z|^2 \;|\; s) + (\alpha - \lambda) \frac{\mathbb{E}\big[s^2\big] + M}{\rho + M} \\
& \quad - h\big(\underline{/s + z} + \Delta \,\big|\, |s + z|\,, s\big) \bigg\} \\
\le{} & \alpha \log \frac{\rho + M}{\alpha} + d_{\lambda,\alpha} + \log(2\pi) \\
& + \max_{\xi \ge 0} g_{\lambda,\alpha}(\xi, \rho)
\end{aligned} \tag{15}
$$

where $g_{\lambda,\alpha}(\xi, \rho)$ was defined in (7). In the last step, we upper-bounded the supremum over $\mathcal{Q}_s$ with the supremum over all deterministic $\xi \ge 0$. The resulting upper bound can be tightened by minimizing it over α and λ, which yields (6). This concludes the proof. ■

It is instructive to note that if we further lower-bound the last term on the RHS of (12) as

$$
\begin{aligned}
& h\big(\underline{/s + z} + \Delta \,\big|\, |s + z|\,, s\big) \\
& \ge h\big(\underline{/s + z} + \Delta \,\big|\, |s + z|\,, s, \underline{/s + z}\big) \\
& = h(\Delta)
\end{aligned}
$$

we obtain

$$
C(\rho) \le \sup_{\mathcal{Q}_s} \big\{ I(s, \mathbf{v}; \mathbf{y}) \big\} + \log(2\pi) - h(\Delta) \tag{16}
$$

where $s\mathbf{v}$ is the input to the memoryless phase-noise channel with uniform phase noise (13). The inequality (16) can be interpreted as follows: the capacity of a Wiener phase-noise channel is upper-bounded by the capacity of a memoryless phase-noise channel with uniform phase noise, plus a correction term that accounts for the memory in the channel and does not depend on the SNR ρ.

If we now specialize (16) to single antenna systems and we add the additional constraint on $\mathcal{Q}_s$ that $|s|^2 = \rho$ with probability one (which holds, for example, if a PSK constellation is used), the first term on the RHS of (16) vanishes and we recover the upper bound previously reported in [9, Theorem 2].

V. Large-SNR Regime

In Theorem 3 below, we present an asymptotic characterization of $C(\rho)$, which generalizes to the MIMO case the asymptotic characterization reported in [4] for the single-antenna case.

Theorem 3: In the large-SNR regime, the capacity of the Wiener phase-noise channel (1) behaves as

$$
\begin{aligned}
C(\rho) = {} & \left(M - \frac{1}{2}\right) \log \frac{2\rho}{2M - 1} + \log \frac{\Gamma(M - 1/2)}{\Gamma(M)} \\
& + \frac{1}{2} \log \pi - h(\Delta) + o(1)
\end{aligned} \tag{17}
$$

where $o(1)$ indicates a function of ρ that vanishes in the limit $\rho \to \infty$.

Proof: The asymptotic characterization (17) is obtained by proving that an appropriately modified version of the upper bound presented in Section IV matches the lower bound we shall report in this section up to a $o(1)$ term.

Upper bound: We exploit the property that the high-SNR behavior of $C(\rho)$ does not change if the support of the input distribution is constrained to lie outside a sphere of arbitrary radius. This result, known as *escape-to-infinity* property of the capacity-achieving input distribution [14, Def. 4.11], is formalized in the following lemma.

Lemma 4: Fix an arbitrary $\xi_0 > 0$ and let $\mathcal{K}(\xi_0) = \{\mathbf{x} \in \mathbb{C}^M : \|\mathbf{x}\| \geq \xi_0\}$. Denote by $C^{(\xi_0)}(\rho)$ the capacity of the channel (1) when the input signal is subject to the average-power constraint (5) and to the additional constraint that $\mathbf{x}_k \in \mathcal{K}(\xi_0)$ almost surely for all k. Then

$$C(\rho) = C^{(\xi_0)}(\rho) + o(1), \quad \rho \to \infty$$

with $C(\rho)$ given in (4).

Proof: The lemma follows directly from [15, Thm. 8] and [14, Thm. 4.12]. ■

Fix $\xi_0 > 0$. By performing the same steps leading to (15), but accounting for the additional constraint that $\mathbf{x} \in \mathcal{K}(\xi_0)$ almost surely and also setting $\alpha = \lambda = M-1/2$, we obtain: $C^{(\xi_0)}(\rho) \leq U^{(\xi_0)}(\rho)$, where

$$U^{(\xi_0)}(\rho) \triangleq \left(M - \frac{1}{2}\right) \log \frac{2(\rho + M)}{2M-1} + \log \frac{\Gamma(M-1/2)}{\Gamma(M)} + \log(2\pi) + \frac{1}{2} + \max_{\xi \geq \xi_0}\{\tilde{g}(\xi)\}. \quad (18)$$

with $\tilde{g}(\xi) \triangleq g_{\lambda,\alpha}(\xi,\rho)\,|_{\lambda=\alpha=M-1/2}$. As $\lim_{\xi\to\infty} \tilde{g}(\xi) = -(1/2)\log(4\pi e) - h(\Delta)$ (see [4, Eq. (9)] and proceed similarly to the proof of [14, Lemma 6.9]), we can make (18) to be arbitrarily close to (17) by choosing ξ_0 sufficiently large.

Lower bound: Take $\{\mathbf{x}_k\}$ i.i.d. and isotropically distributed with

$$\|\mathbf{x}_k\|^2 = \rho \frac{t_{\rho,\rho_0}}{(M-1/2)} \quad (19)$$

where, for a given $\rho_0 > 0$, the random variable t_{ρ,ρ_0} has pdf

$$f_{t_{\rho,\rho_0}}(a) = \begin{cases} \dfrac{f^{(\rho,\rho_0)}(a)}{\Pr\{z > \rho_0/\rho\}}, & \text{if } a > \rho_0/\rho \\ 0, & \text{otherwise.} \end{cases}$$

Here, $f^{(\rho,\rho_0)}$ denotes the pdf of a random variable that follows a $\text{Gamma}((M-1/2)\cdot\Pr\{z > \rho_0/\rho\}, 1)$ distribution. Let $t \sim \text{Gamma}(M-1/2, 1)$ and denote its pdf by f_t. Note that for all ρ_0 the pdf $f_{t_{\rho,\rho_0}}$ converges point-wise to f_t as $\rho \to \infty$. As

$$\mathbb{E}[t_{\rho,\rho_0}] \leq M - 1/2$$

the average-power constraint (5) is satisfied. A key feature of the distribution of $\|\mathbf{x}_k\|$ in (19) is that $\Pr\{\|\mathbf{x}_k\| < \xi_0\} = 0$ where

$$\xi_0^2 \triangleq \frac{\rho_0}{M-1/2}. \quad (20)$$

Note that $\xi_0 \to \infty$ as $\rho_0 \to \infty$, a property that will be useful in the remainder of the proof.

To obtain the desired lower bound, we use chain rule for mutual information and that mutual information is nonnegative

$$\begin{aligned} I(\mathbf{x}^n; \mathbf{y}^n) &= \sum_{k=1}^n I(\mathbf{x}_k; \mathbf{y}^n \,|\, \mathbf{x}^{k-1}) \\ &\geq \sum_{k=2}^n I(\mathbf{x}_k; \mathbf{y}^k \,|\, \mathbf{x}^{k-1}). \end{aligned} \quad (21)$$

Fix now $k \geq 2$ and set

$$\epsilon_k \triangleq I(\mathbf{x}_k; \theta_{k-1} \,|\, \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}).$$

We have

$$\begin{aligned} I(\mathbf{x}_k; \mathbf{y}^k \,|\, \mathbf{x}^{k-1}) &\overset{(a)}{=} I(\mathbf{x}_k; \mathbf{y}^k, \mathbf{x}^{k-1}) \\ &\overset{(b)}{\geq} I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}) \\ &= I(\mathbf{x}_k; \mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{x}_{k-1}, \theta_{k-1}) - \epsilon_k \\ &\overset{(c)}{=} I(\mathbf{x}_k; \mathbf{y}_k, \theta_{k-1}) - \epsilon_k \\ &\overset{(d)}{=} I(\mathbf{x}_k; \mathbf{y}_k \,|\, \theta_{k-1}) - \epsilon_k \\ &\overset{(e)}{=} I(\mathbf{x}_2; \mathbf{y}_2 \,|\, \theta_1) - \epsilon_2. \end{aligned} \quad (22)$$

Here, (a) follows because $\{\mathbf{x}_k\}$ are independent; in (b) we used chain rule for mutual information and that mutual information is nonnegative; (c) follows because $\mathbf{x}_k$ and the pair $(\mathbf{y}_{k-1}, \mathbf{x}_{k-1})$ are conditionally independent given $(\theta_{k-1}, \mathbf{y}_k)$; (d) holds because $\mathbf{x}_k$ and θ_{k-1} are independent; finally (e) follows from stationarity. Substituting (22) into (21), we obtain

$$C(\rho) \geq I(\mathbf{x}_2; \mathbf{y}_2 \,|\, \theta_1) - \epsilon_2. \quad (23)$$

We next investigate the two terms on the RHS of (23) separately. Specifically, we shall show that the first term has the desired asymptotic expansion, while the second term can be made arbitrarily close to zero by choosing ρ_0 in (19) sufficiently large.

A. The first term on the RHS of (23)

We write

$$I(\mathbf{x}_2; \mathbf{y}_2 \,|\, \theta_1) = h(\mathbf{y}_2 \,|\, \theta_1) - h(\mathbf{y}_2 \,|\, \mathbf{x}_2, \theta_1) \quad (24)$$

and bound the two terms separately. For the first term, we have that

$$\begin{aligned} h(\mathbf{y}_2 \,|\, \theta_1) &\geq h(\mathbf{y}_2 \,|\, \mathbf{w}_2, \theta_1) \\ &= h(e^{j\theta_2}\mathbf{x}_2 \,|\, \theta_1) \\ &\overset{(a)}{=} h(\mathbf{x}_2) \\ &\overset{(b)}{=} h(\|\mathbf{x}_2\|^2) + \log \frac{\pi^M}{\Gamma(M)} + (M-1)\,\mathbb{E}\left[\log\|\mathbf{x}_2\|^2\right] \\ &\overset{(c)}{=} M \log \frac{\rho}{M-1/2} + \log \frac{\pi^M}{\Gamma(M)} \\ &\quad + h(t_{\rho,\rho_0}) + (M-1)\,\mathbb{E}[\log t_{\rho,\rho_0}]\,. \end{aligned}$$

Here, (a) follows because $\mathbf{x}_2$ is isotropically distributed, and, hence, $e^{j\theta_2}\mathbf{x}_2 \sim \mathbf{x}_2$; in (b) we computed the differential entropy in polar coordinates [14, Lemma 6.15 and 6.17]; finally, (c) follows from (19). For the second term on the RHS of (24), we proceed as follows. Let $\mathbf{x}_2 = s_2\mathbf{v}_2$, with $s_2 = \|\mathbf{x}_2\|$ and, hence, $s_2^2 \sim \rho\, t_{\rho,\rho_0}/(M-1/2)$. Furthermore, let $z_2 \sim \mathcal{CN}(0,1)$. Then

$$\begin{aligned} h(\mathbf{y}_2 \,|\, \mathbf{x}_2, \theta_1) &= h(\mathbf{y}_2 \,|\, s_2, \mathbf{v}_2, \theta_1) \\ &= h(e^{j\theta_2} s_2 + z_2 \,|\, s_2, \theta_1) + \log(\pi e)^{M-1}. \end{aligned}$$

Now note that

$$\begin{aligned} & h(e^{j\theta_2} s_2 + z_2 \,|\, s_2, \theta_1) \\ &\;= h(e^{j\theta_2}(s_2 + z_2) \,|\, s_2, \theta_1) \\ &\overset{(a)}{=} h(e^{j\Delta}(s_2 + z_2) \,|\, s_2) \\ &\overset{(b)}{=} h(|s_2 + z_2|^2 \;|\; s_2) \\ &\quad + h(\angle s_2 + z_2 + \Delta \;\big|\; |s_2 + z_2|, s_2) - \log 2 \\ &\overset{(c)}{\leq} \frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1 + \frac{4\rho}{2M-1} t_{\rho,\rho_0}\right]\right)\right] \\ &\quad + h(\Delta + \angle s_2 + z_2 \,|\, s_2) - \log 2. \end{aligned}$$

Here, in (a) we used (2) and denoted by Δ a random variable distributed as in (3); in (b) we evaluated the differential entropy in polar coordinates [14, Lemma 6.15 and 6.16]. Finally, (c) follows because the Gaussian distribution maximizes differential entropy under a variance constraint and because conditioning reduces entropy. Note finally that

$$\begin{aligned} h(\angle s_2 + z_2 + \Delta \,|\, s_2) &\leq \max_{\xi \geq \xi_0} h(\angle \xi + z_2 + \Delta) \\ &= h(\angle \xi_0 + z_2 + \Delta). \end{aligned}$$

This term can be made arbitrarily close to $h(\Delta)$ by choosing ρ_0 in (20) sufficiently large. Summarizing, we have shown that

$$\begin{aligned} I(\mathbf{x}_2; \mathbf{y}_2 \,|\, \theta_1) &\geq M \log \frac{2\rho}{2M-1} + \log \frac{\pi^M}{\Gamma(M)} + h(t_{\rho,\rho_0}) \\ &\quad + (M-1)\,\mathbb{E}[\log t_{\rho,\rho_0}] \\ &\quad - \frac{1}{2}\mathbb{E}\left[\log\left(2\pi e\left[1 + \frac{4\rho}{2M-1} t_{\rho,\rho_0}\right]\right)\right] \\ &\quad - h(\angle \xi_0 + z_2 + \Delta) - \log\left[2(\pi e)^{M-1}\right] \\ &\overset{(a)}{=} \left(M - \frac{1}{2}\right) \log \frac{2\rho}{2M-1} + \log \frac{\Gamma(M-1/2)}{\Gamma(M)} \\ &\quad + \frac{1}{2}\log \pi - h(\angle \xi_0 + z_2 + \Delta) + o(1). \end{aligned}$$

Here, (a) follows because

$$\begin{aligned} h(t_{\rho,\rho_0}) &= h(t) + o(1) \\ \mathbb{E}[\log(t_{\rho,\rho_0})] &= \mathbb{E}[\log t] + o(1) \\ \mathbb{E}[\log(1 + c\rho t_{\rho,\rho_0})] &= \log(c\rho) + \mathbb{E}[\log t] + o(1), \forall c > 0 \end{aligned}$$

where $t \sim \text{Gamma}(M - 1/2, 1)$ and because

$$\begin{aligned} \mathbb{E}[\log t] &= \psi(M - 1/2) \\ h(t) &= (3/2 - M)\psi(M - 1/2) \\ &\quad + M - 1/2 + \log \Gamma(M - 1/2) \end{aligned}$$

with $\psi(\cdot)$ denoting Euler's digamma function.

B. The second term on the RHS of (23)

Let $\mathbf{x}_1 = s_1\mathbf{v}_1$ and $z_1 \sim \mathcal{CN}(0,1)$. Proceeding similarly as in [14, App. IX], we obtain

$$\begin{aligned} & I(\mathbf{x}_2; \theta_1 \,|\, \mathbf{y}_2, \mathbf{y}_1, \mathbf{x}_1) \\ &= h(\theta_1 \,|\, \mathbf{y}_2, \mathbf{y}_1, \mathbf{x}_1) - h(\theta_1 \,|\, \mathbf{y}_2, \mathbf{x}_2, \mathbf{y}_1, \mathbf{x}_1) \\ &\leq h(\theta_1 \,|\, \mathbf{y}_1, \mathbf{x}_1) - h(\theta_1 \,|\, \mathbf{y}_2, \mathbf{x}_2, \mathbf{y}_1, \mathbf{x}_1, \theta_2) \\ &= h(\theta_1 \,|\, \mathbf{y}_1, \mathbf{x}_1) - h(\theta_1 \,|\, \mathbf{y}_1, \mathbf{x}_1, \theta_2) \\ &= I(\theta_1; \theta_2 \,|\, \mathbf{y}_1, \mathbf{x}_1) \\ &= h(\theta_2 \,|\, \mathbf{y}_1, \mathbf{x}_1) - h(\theta_2 \,|\, \mathbf{y}_1, \mathbf{x}_1, \theta_1) \\ &= h(\theta_2 \,|\, e^{j\theta_1}(s_1 + z_1), s_1) - h(\theta_2 \,|\, \theta_1) \\ &\leq \max_{\xi \geq \xi_0} h(\theta_2 \,|\, e^{j\theta_1}(\xi + z_1)) - h(\theta_2 \,|\, \theta_1) \\ &= h(\theta_2 \,|\, e^{j\theta_1}(\xi_0 + z_1)) - h(\theta_2 \,|\, \theta_1). \end{aligned} \tag{25}$$

As claimed, the RHS of (25) can be made arbitrarily close to zero by choosing ρ_0 in (20) sufficiently large. ■

VI. Simulation Results

In this section, we numerically evaluate the upper bound $U(\rho)$ in (6) and the asymptotic capacity expression (17)—the $o(1)$ term is neglected—for a single-antenna Wiener phase-noise channel with standard deviation of the phase-noise increment equal to 6° (Fig. 1) and 20° (Fig. 2). In the figures, we also show the capacity

$$C_{\text{awgn}}(\rho) \triangleq \log(1 + \rho) \tag{26}$$

of an AWGN channel with SNR equal to ρ, which is a tight upper bound on $C(\rho)$ at low SNR. We also display

$$\tilde{U}(\rho) \triangleq \min\{C_{\text{awgn}}(\rho), U(\rho)\}. \tag{27}$$

The information rates achievable using QAM constellations of different cardinality, which are lower bounds on $C(\rho)$, are also depicted. As in [9], [2], [10], we evaluate these rates using the algorithm for the computation of the information rates for finite-state channels proposed in [11]. Specifically, we use 200 levels for the discretization of the phase-noise process, and a block of 2000 channel uses.

In both scenarios C_{awgn} is a tighter upper bound than $U(\rho)$ at low SNR values where the additive noise is the main impairment. In the high-SNR regime, however, $U(\rho)$ is tighter. For medium/high SNR values, the large-SNR capacity approximation (17) follows $U(\rho)$ closely. In this regime, QAM constellations incur a penalty of more than 3 dB. The gap to the capacity upper bound might be reduced by replacing QAM with suitably optimized constellations.

Appendix

We follow an approach similar to the one pursued in [8] for the single-antenna block-constant phase-noise case. Let $q_{\mathbf{y}}(\mathbf{y})$ denote an arbitrary pdf on $\mathbf{y}$. By duality [14, Thm. 5.1], for every probability distribution Q_s on s we have that

$$I(s, \mathbf{v}; \mathbf{y}) \leq -\mathbb{E}[\log q_{\mathbf{y}}(\mathbf{y})] - h(\mathbf{y} \,|\, s, \mathbf{v}). \tag{28}$$

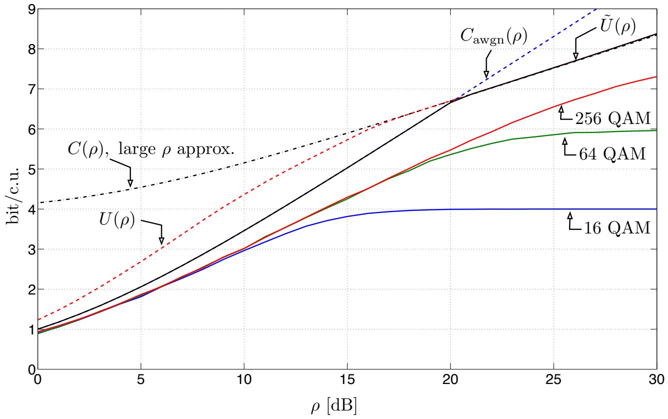

Fig. 1. The upper bound $U(\rho)$ in (6), the asymptotic capacity approximation (17), the AWGN capacity (26), the tighter upper bound $\tilde{U}(\rho)$ in (27), and the rates achievable with 16, 64, and 256 QAM. In the figure, $\sigma_\Delta = 6°$.

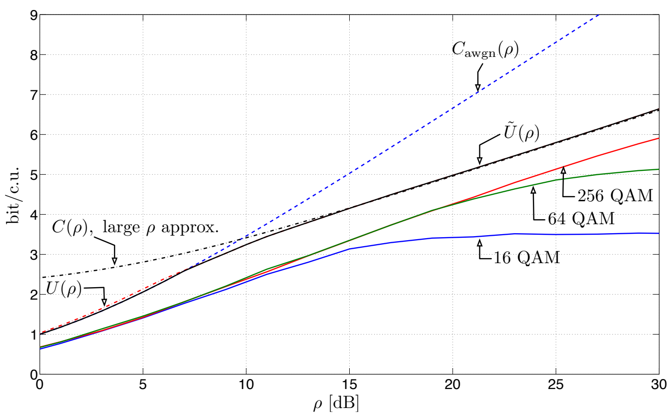

Fig. 2. The upper bound $U(\rho)$ in (6), the asymptotic capacity approximation (17), the AWGN capacity (26), the tighter upper bound $\tilde{U}(\rho)$ in (27), and the rates achievable with 16, 64, and 256 QAM. In the figure, $\sigma_\Delta = 20°$.

The expectation on the RHS of (28) is with respect to the probability distribution induced on $\mathbf{y}$ by the distribution on s (which we need to optimize over) and the uniform distribution on $\mathbf{v}$ through (13). Note also that for every probability distribution $\mathcal{Q}_s$ satisfying $\mathbb{E}\left[s^2\right] \le \rho$, we have that

$$1 - \left[\mathbb{E}\left[s^2\right] + M\right] / \left[(\rho + M)\right] \ge 0. \tag{29}$$

Fix now $\lambda \ge 0$ and an arbitrary pdf $q_{\mathbf{y}}(\mathbf{y})$ on $\mathbf{y}$. Using (28) and (29), we conclude that for every probability distribution $\mathcal{Q}_s$ on s

$$\begin{aligned} I(s, \mathbf{v}; \mathbf{y}) \le &-\mathbb{E}[\log q_{\mathbf{y}}(\mathbf{y})] - h(\mathbf{y} \,|\, s, \mathbf{v}) \\ &+ \lambda \left(1 - \frac{\mathbb{E}\left[s^2\right] + M}{\rho + M}\right). \end{aligned} \tag{30}$$

Let $\mathbf{y} = \sqrt{r} \cdot \mathbf{u}$, where $r = \|\mathbf{y}\|^2$ and $\mathbf{u} = \mathbf{y}/\|\mathbf{y}\|$. To evaluate the first term on the RHS of (30), we take $q_{\mathbf{y}}(\mathbf{y})$ so that

$$q_r(r) = \frac{r^{\alpha-1} e^{-r/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad r \ge 0 \tag{31}$$

with α to be optimized later, and $\beta = (\rho + M)/\alpha$. Furthermore, we take $\mathbf{u}$ uniformly distributed on the unit sphere in $\mathbb{C}^M$ and independent of r. By using polar coordinates,

$$\begin{aligned} -\mathbb{E}[\log q_{\mathbf{y}}(\mathbf{y})] &= -\mathbb{E}\left[\log q_{\mathbf{y}}(\sqrt{r} \cdot \mathbf{u})\right] \\ &\overset{(a)}{=} -\mathbb{E}[\log q_r(r)] + \log \frac{\pi^M}{\Gamma(M)} \\ &\quad + (M-1)\,\mathbb{E}[\log r] \\ &\overset{(b)}{=} (M - \alpha)\,\mathbb{E}[\log r] + \alpha \frac{\mathbb{E}[r]}{\rho + M} \\ &\quad + \log\left(\pi^M \frac{\Gamma(\alpha)}{\Gamma(M)}\right) \\ &\quad + \alpha \log \frac{\rho + M}{\alpha}. \end{aligned} \tag{32}$$

Here, in (a) we used that

$$q_{r,\mathbf{u}}(r, \mathbf{u}) = q_{\mathbf{y}}\left(\sqrt{r} \cdot \mathbf{u}\right) \cdot r^{M-1}/2$$

as a consequence of the change of variable theorem, and that

$$q_{r,\mathbf{u}}(r, \mathbf{u}) = q_r(r) \cdot \Gamma(M)/(2\pi^M)$$

by construction; (b) follows from (31) with $\beta = (\rho + M)/\alpha$. We next compute the conditional differential entropy term in (30). Let $z \sim \mathcal{CN}(0, 1)$. We have that

$$\begin{aligned} h(\mathbf{y} \,|\, s, \mathbf{v}) &= h(se^{j\theta} + z \,|\, s) + \log(\pi e)^{M-1} \\ &= h(e^{j\theta}(s + z) \,|\, s) + \log(\pi e)^{M-1} \\ &= h(|s + z|^2 \;|\; s) + \log \pi^M + M - 1. \end{aligned} \tag{33}$$

The last step follows by using [14, Lem. 6.16] and that $\theta \sim \mathcal{U}[0, 2\pi]$. Let $d_{\lambda,\alpha}$ be defined as in (8). By substituting (32) and (33) into (30) and by using that

$$\mathbb{E}[r] = \mathbb{E}\left[s^2\right] + M$$

and that

$$\mathbb{E}[\log r] = \mathbb{E}\left[\log\left(|s + z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right]$$

where $z_1, \ldots, z_M$ are independent $\mathcal{CN}(0, 1)$-distributed random variables, we obtain

$$\begin{aligned} I(s, \mathbf{v}; \mathbf{y}) \le \alpha \log \frac{\rho + M}{\alpha} &+ d_{\lambda,\alpha} \\ &+ (M - \alpha)\,\mathbb{E}\left[\log\left(|s + z_1|^2 + \sum_{j=2}^{M} |z_j|^2\right)\right] \\ &- h(|s + z|^2 \;|\; s) + (\alpha - \lambda) \frac{\mathbb{E}\left[s^2\right] + M}{\rho + M}. \end{aligned}$$

References

[1] G. Durisi, A. Tarable, and T. Koch, "On the multiplexing gain of MIMO microwave backhaul links affected by phase noise," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, to appear.

[2] G. Colavolpe, "Communications over phase-noise channels: A tutorial review," in *Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), 2012 6th*, Sep. 2012, pp. 316–327.

[3] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Effect of oscillator phase noise on uplink performance of large MU-MIMO systems," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, U.S.A., Oct. 2012.

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

[5] H. Ghozlan and G. Kramer, "On Wiener phase noise channels at high signal-to-noise ratio," submitted to IEEE Int. Symp. Inf. Theory (ISIT), Jan. 2013. [Online]. Available: http://arxiv.org/abs/1301.6923

[6] R. Nuriyev and A. Anastasopoulos, "Capacity and coding for the block-independent noncoherent AWGN channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 866–883, Mar. 2005.

[7] 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.

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

[9] A. Barbieri and G. Colavolpe, "On the information rate and repeat-accumulate code design for phase noise channels," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3223 –3228, Dec. 2011.

[10] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time wiener's phase noise channel," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1480–1486, May 2012.

[11] 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.

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

[13] 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.

[14] 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.

[15] ——, "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.