Chalmers Publication Library

Copyright Notice

(Article begins on next page)

On the Multiplexing Gain of MIMO Microwave Backhaul Links Affected by Phase Noise

Giuseppe Durisi
Chalmers University of Technology
41296 Gothenburg, Sweden
E-mail: durisi@chalmers.se

Alberto Tarable
IEIIT-CNR
10129 Turin, Italy
E-mail: alberto.tarable@ieiit.cnr.it

Tobias Koch
Universidad Carlos III de Madrid
28911 Leganés, Spain
E-mail: koch@tsc.uc3m.es

Abstract—We consider a multiple-input multiple-output (MIMO) AWGN channel affected by phase noise. Focusing on the 2×2 case, we show that no MIMO multiplexing gain is to be expected when the phase-noise processes at each antenna are independent, memoryless in time, and with uniform marginal distribution over $[0, 2\pi]$ (strong phase noise), and when the transmit signal is isotropically distributed on the real plane. The scenario of independent phase-noise processes across antennas is relevant for microwave backhaul links operating in the 20–40 GHz range.

I. INTRODUCTION

One common solution to the problem of guaranteeing backhaul connectivity in mobile cellular networks is to use point-to-point microwave links. The current terrific rate of increase in mobile data traffic makes these microwave radio links a potential bottleneck in the deployment of high-throughput cellular networks. This consideration has stimulated a large body of research aimed at designing high-capacity backhaul links [1]. One design challenge is that the use of high-order constellations (512 QAM has been recently demonstrated in commercial products) to increase throughput makes the overall system extremely sensitive to *phase noise*, i.e., to phase and frequency instabilities in the radio-frequency (RF) oscillators used at the transmitter and the receiver.

The use of multiple antennas is currently investigated as a solution to increase the throughput of microwave backhaul links. These links are typically in line of sight (LOS). Hence, neglecting phase noise, to release multiple-input multiple-output (MIMO) multiplexing gains—i.e., a throughput increase proportional to the minimum between the number of transmit and of receive antennas—the antennas at the transmitter and at the receiver must be spaced sufficiently far apart. For example, for backhaul microwave links operating in the commonly used 20–40 GHz range, the spacing between antennas must be on the order of few meters.

In such a setup, cost considerations imply that the RF circuitries connected to each antenna are driven by independent oscillators. This, in turn, yields independent phase-noise processes at each antenna. As the number of independent phase-noise processes that need to be tracked at the receiver increases with the number of antennas, it is unclear whether MIMO multiplexing

gains can be actually expected in this setup. In this paper, we will indeed show that if the phase-noise process is uniformly distributed over $[0, 2\pi]$ and memoryless in time, and if it is independent across antennas, then no MIMO multiplexing gains can be expected when the direction of the transmit signal is isotropically distributed. In the rest of the paper, we shall refer to phase-noise processes that are uniformly distributed over $[0, 2\pi]$ and are memoryless in time as *strong phase noise*.

Before detailing the contributions of this paper, we present a brief review of the results available in the literature concerning the capacity of phase-noise channels.

A Brief Literature Review: The capacity of the phase-noise channel is not known in closed form, even for the single-antenna case. Capacity bounds for strong phase noise have been reported in [2]. These bounds, which turn out to be tight over a large range of signal-to-noise ratio (SNR) values, have been recently extended to the case of block-memoryless uniformly distributed phase noise in [3]. The rates achievable with Gaussian inputs on a Wiener-distributed phase-noise channel have been characterized in [4].

A high-SNR capacity characterization that holds for the general class of stationary phase-noise processes with finite differential-entropy rate has been obtained in [5]. Roughly speaking, the result in [5] implies that whenever the phase-noise process satisfies the property that its current realizations cannot be perfectly predicted from the observation of the process' infinite past, then—at high SNR—the capacity of the corresponding phase-noise channel is half the capacity of an AWGN channel with the same receive SNR. In mathematical terms, the *pre-log*, i.e., the asymptotic ratio between the capacity and the logarithm of SNR as SNR goes to infinity, of every single-antenna phase-noise channel belonging to this class is $1/2$. Although the impact of phase noise in the measurement of MIMO channels has been addressed in the literature [6], no capacity characterizations of the same nature as the one reported in [5] are available for MIMO phase-noise channels, to the best of the authors' knowledge.

Receiver architectures for the single-antenna case based on joint phase-noise recovery and decoding have been proposed, e.g., in [7]–[9]. An extension of some of these architectures to the MIMO case can be found in [10].

Contributions: We analyze the capacity of a MIMO system equipped with 2 transmit and 2 receive antennas and operating over an AWGN channel impaired by strong phase noise, namely,

uniformly distributed phase noise that is memoryless over time. Our contributions are as follows:

- When the phase-noise processes are perfectly correlated at the transmitter side and independent at the receiver side (or vice-versa), we show that the capacity pre-log is 1. In this scenario, MIMO yields the expected 2-fold capacity increase at high SNR. This result can be extended to the general $N \times N$ MIMO case: the capacity pre-log is, in this case, equal to $N/2$.
- When the phase-noise processes are perfectly correlated both at the transmitter side and at the receiver side, i.e., when the RF circuitries are driven by the same oscillator, we show that the capacity pre-log for the $N \times N$ MIMO case is even larger and equals $N - 1/2$. This setup is relevant for microwave backhaul links operating in the high frequency range (70 GHz and above), for which the antenna spacing is small enough to allow for a single oscillator.
- When the phase-noise processes are independent across antennas (microwave backhaul links operating in the 20–40 GHz range), we provide a characterization of the rates achievable under the constraint that the direction of the input signal is isotropically distributed on the real plane and independent of its magnitude. A distribution that satisfies this property achieves the high-SNR capacity when the phase-noise processes are perfectly correlated at either the transmitter or the receiver side. We show that, in this case, the pre-log is $1/2$ (same pre-log as for a single-antenna phase-noise channel). This means that in this case MIMO yields no multiplexing gain.

Notation

Boldface letters (e.g., $\mathbf{a}$) denote random quantities, while the ordinary font (e.g., a) is reserved for their realizations and for deterministic quantities. Underlined uppercase letters are used to indicate matrices (for example, $\underline{\mathbf{A}}$ is a random matrix and $\underline{A}$ denotes its realizations), while underlined lowercase letters are reserved for vectors. Uppercase sans-serif letters (e.g., P) denote probability distributions, while lowercase sans-serif letters (e.g., q) are reserved for probability density functions (pdf). If two random vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ have the same distribution, we write $\underline{\mathbf{a}} \sim \underline{\mathbf{b}}$. The relative entropy between the two distributions $\mathsf{P}_{\mathbf{x}}$ and $\mathsf{P}_{\mathbf{y}}$ is denoted by $D(\mathsf{P}_{\mathbf{x}} \| \mathsf{P}_{\mathbf{y}})$.

The superscripts ^{T} and ^{H} stand for transposition and Hermitian transposition, respectively. We denote the identity matrix of dimension $N \times N$ by $\underline{\mathbf{I}}_N$; $\mathrm{diag}\{\underline{a}\}$ is the diagonal square matrix whose main diagonal contains the entries of the vector $\underline{a}$. For a given vector $\underline{a}$, we denote by $|\underline{a}|$ the vector whose entries are the absolute value of the entries of $\underline{a}$, i.e., $[|\underline{a}|]_i = |[\underline{a}]_i|, \forall i$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = \mathcal{O}(g(x))$, $x \to \infty$, means that $\limsup_{x \to \infty} |f(x)/g(x)| < \infty$, and $f(x) = o(g(x))$, $x \to \infty$, means that $\lim_{x \to \infty} |f(x)/g(x)| = 0$. We say that a random variable $\mathbf{r}$ has Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ and write $\mathbf{r} \sim \mathrm{Gamma}(\alpha, \beta)$ if

its pdf $\mathsf{q}_{\mathbf{r}}(r)$ is given by

$$\mathsf{q}_{\mathbf{r}}(r) = \frac{r^{\alpha-1} e^{-r/\beta}}{\beta^{\alpha} \Gamma(\alpha)}, \quad r \geq 0. \quad (1)$$

Here, $\Gamma(\cdot)$ denotes the Gamma function. We shall use the property that if $\mathbf{r} \sim \mathrm{Gamma}(\alpha, \beta)$, then $\mathbb{E}[\mathbf{r}] = \alpha\beta$. Finally, $\log(\cdot)$ indicates the natural logarithm.

II. SYSTEM MODEL

We consider the following $N \times N$ MIMO phase-noise channel

$$\underline{\mathbf{y}} = \underline{\mathbf{\Phi}}_R \, \underline{\mathbf{H}} \, \underline{\mathbf{\Phi}}_T \underline{x} + \underline{\mathbf{w}}. \quad (2)$$

Here, $\underline{x}$ and $\underline{\mathbf{y}}$ are N-dimensional (complex-valued) vectors that contain the transmitted symbols and received samples, respectively; $\underline{\mathbf{H}}$ is the $N \times N$ channel matrix, which is assumed to be deterministic (MIMO LOS scenario) and known both at the transmitter and at the receiver; $\underline{\mathbf{\Phi}}_T = \mathrm{diag}\left([e^{j\phi_1^t} \cdots e^{j\phi_N^t}]\right)$ and $\underline{\mathbf{\Phi}}_R = \mathrm{diag}\left([e^{j\phi_1^r} \cdots e^{j\phi_N^r}]\right)$ are *independent* diagonal matrices containing the phase-noise samples at the transmitter and at the receiver, respectively; and $\underline{\mathbf{w}}$ is an N-dimensional (complex-valued) vector containing uncorrelated, circularly-symmetric Gaussian noise samples, with zero mean and variance $1/2$ per real dimension.

We assume that neither the transmitter nor the receiver are aware of the realizations of $\underline{\mathbf{\Phi}}_T$ and $\underline{\mathbf{\Phi}}_R$. We shall further assume that the channel matrix $\underline{\mathbf{H}}$ is unitary, i.e.,

$$\underline{\mathbf{H}}^H \underline{\mathbf{H}} = \underline{\mathbf{I}}_N.$$

This condition holds in the MIMO LOS scenario, where the channel coefficients are controllable by properly adjusting the distance between antennas [11]. We shall finally assume that the phase-noise samples $\{\phi_n^t\}_{n=1}^N$ and $\{\phi_n^r\}_{n=1}^N$ are uniformly distributed over $[0, 2\pi]$ and that they take on independent values at every channel use. This assumption corresponds to the worst-case situation of strong phase noise. In Section IV, we discuss how to account for situations where the channel matrix $\underline{\mathbf{H}}$ is not unitary and the phase-noise process has memory.

We shall consider the following two scenarios:

- Both $\underline{\mathbf{\Phi}}_T$ and $\underline{\mathbf{\Phi}}_R$ have independent entries. This corresponds to the case of independent oscillators per antenna.
- $\phi_1^t = \cdots = \phi_N^t$ and (or) $\phi_1^r = \cdots = \phi_N^r$. This corresponds to the case of a single oscillator at the transmitter side, and (or) at the receiver side.

III. CAPACITY AND PRE-LOG

A. *Definitions*

The capacity of the channel in (2) is given by [12]

$$C(\rho) = \sup_{\mathsf{P}_{\underline{\mathbf{x}}}} I(\underline{\mathbf{x}}; \underline{\mathbf{y}}) \quad (3)$$

where the supremum is over the set of probability distributions $\mathsf{P}_{\underline{\mathbf{x}}}$ that satisfy the average-power constraint

$$\mathbb{E}\left[\|\underline{\mathbf{x}}\|^2\right] \leq \rho. \quad (4)$$

As the channel matrix $\underline{H}$ is unitary, and as the variance of the additive noise is normalized to 1, we can interpret ρ in (4) as the receive SNR.

The goal of this paper is to characterize the capacity pre-log χ, which is defined as

$$\chi = \limsup_{\rho \to \infty} \frac{C(\rho)}{\log(\rho)}.$$

For the single-antenna case (i.e., $N = 1$), Lapidoth showed that [5]

$$C(\rho) = \frac{1}{2}\log(\rho) - \frac{1}{2}\log(2) + o(1), \quad \rho \to \infty \quad (5)$$

which implies that the pre-log for the single-antenna case is $1/2$, i.e., half of that of an AWGN channel with the same receive power. It is worth recalling that the probability distribution on the *scalar* input $\mathbf{x}$ that achieves the high-SNR capacity (5) is such that $|\mathbf{x}|^2 \sim \text{Gamma}(1/2, 2\rho)$. The phase of $\mathbf{x}$ cannot be used to transmit information.

B. The Parallel Phase-Noise Channel

We next state a simple consequence of (5), which will be useful in our analysis.

Proposition 1: The capacity of the $N \times N$ memoryless parallel phase-noise channel

$$\underline{y} = \underline{\Phi}\,\underline{x} + \underline{w} \quad (6)$$

with $\underline{\Phi} = \text{diag}\big([e^{j\phi_1} \cdots e^{j\phi_N}]\big)$ and $\{\phi_n\}_{n=1}^N$ independent and uniformly distributed over $[0, 2\pi]$ is given by

$$C(\rho) = \frac{N}{2}\log(\rho) - \frac{N}{2}\log(2N) + o(1), \quad \rho \to \infty. \quad (7)$$

Proof: Since $\{\phi_n\}_{n=1}^N$ are independent and uniformly distributed over $[0, 2\pi]$, capacity is achieved by an input distribution for which the entries of $\underline{x}$ are independent and identically distributed. Furthermore, (5) implies that the choice $|\mathbf{x}_n|^2 \sim \text{Gamma}(1/2, 2\rho/N)$ for $n = 1, \ldots, N$ achieves capacity up to a $o(1)$-term. As in the single-antenna case, the phase of the entries of $\underline{x}$ cannot be used to transmit information. ∎

C. Remarks on the 2×2 Case

Focusing on the 2×2 case, we next recall a property of the input distribution that achieves the high-SNR capacity (7). This property will motivate our analysis in Section III-F.

Lemma 2: Let $N = 2$ and let $|\underline{\mathbf{x}}| = \|\underline{\mathbf{x}}\| \cdot |\hat{\underline{\mathbf{x}}}|$ where

$$|\hat{\underline{\mathbf{x}}}| = |\underline{\mathbf{x}}|/\|\underline{\mathbf{x}}\|$$
$$= [\cos(\boldsymbol{\alpha}_x)\ \sin(\boldsymbol{\alpha}_x)]^T$$

for some random angle $\boldsymbol{\alpha}_x$ supported on $[0, \pi/2]$. The high-SNR asymptotic capacity (7) is achieved by an input distribution for which $\boldsymbol{\alpha}_x$ and $\|\underline{\mathbf{x}}\|^2$ are independent, $\boldsymbol{\alpha}_x$ is uniformly distributed over $[0, \pi/2]$, and $\|\underline{\mathbf{x}}\|^2$ is exponentially distributed, i.e.,

$$q_{\|\underline{\mathbf{x}}\|^2}(z) = \frac{1}{\rho}e^{-z/\rho}.$$

Proof: Recall that (7) (for the case $N = 2$) is achieved by an input distribution for which $|\mathbf{x}_n|^2 \sim \text{Gamma}(1/2, \rho)$,

$n = 1, 2$, with $|\mathbf{x}_1|^2$ independent of $|\mathbf{x}_2|^2$. Further note that $|\mathbf{x}_1| = \|\underline{\mathbf{x}}\|\cos(\boldsymbol{\alpha}_x)$ and $|\mathbf{x}_2| = \|\mathbf{x}\|\sin(\boldsymbol{\alpha}_x)$. By the change-of-variable theorem, the joint pdf of $\|\underline{\mathbf{x}}\|^2$ and $\boldsymbol{\alpha}_x$ is

$$\begin{aligned}
q_{\|\underline{\mathbf{x}}\|^2, \alpha_x}(\|\underline{x}\|^2, \alpha_x) &= q_{|\mathbf{x}_1|^2}(\|\underline{x}\|^2\cos^2(\alpha_x)) \\
&\quad \cdot q_{|\mathbf{x}_2|^2}(\|\underline{x}\|^2\sin^2(\alpha_x)) \\
&\quad \cdot 2\|\underline{x}\|^2\cos(\alpha_x)\sin(\alpha_x) \\
&= \frac{2}{\pi\rho}e^{-\|\underline{x}\|^2/\rho}
\end{aligned}$$

where the last step follows by (1) (with $\alpha = 1/2$ and $\beta = \rho$) and by using that $\Gamma(1/2) = \sqrt{\pi}$. ∎

Lemma 2 provides an intuitive explanation of why the pre-log of the 2×2 memoryless parallel phase-noise channel is twice as large as the pre-log of the single-antenna case. At high SNR, both the modulus $\|\underline{\mathbf{x}}\|$ of the transmit vector $|\underline{\mathbf{x}}|$ and its direction $\boldsymbol{\alpha}_x$ can be estimated reliably from $\underline{y}$. Hence, capacity at high SNR is twice as large compared to the single-antenna case, where information can be conveyed only through the modulus of the (scalar) input signal. Furthermore, since no direction in space should be preferred, the uniform distribution for $\boldsymbol{\alpha}_x$ is capacity-achieving at high SNR.

As a side remark, we note that when phase noise in (6) is absent (i.e., $\underline{\Phi} = \underline{I}_2$), the capacity pre-log is 2 and is achieved by choosing $\mathbf{x}_1$ and $\mathbf{x}_2$ complex Gaussian and independent. This, in turn, implies that $\|\underline{\mathbf{x}}\|^2$ follows a chi-square distribution with 4 degrees of freedom (apart from a scaling factor) and that $\hat{\underline{\mathbf{x}}} = \underline{\mathbf{x}}/\|\underline{\mathbf{x}}\|$ is isotropically distributed in $\mathbb{C}^2$, i.e.,

$$\hat{\underline{\mathbf{x}}} = \begin{bmatrix} e^{j\boldsymbol{\theta}_1} & 0 \\ 0 & e^{j\boldsymbol{\theta}_2} \end{bmatrix} \begin{bmatrix} \cos(\boldsymbol{\alpha}_x) \\ \sin(\boldsymbol{\alpha}_x) \end{bmatrix}$$

with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ independent and uniformly distributed over $[0, 2\pi]$, and $\boldsymbol{\alpha}_x$ such that $\cos^2(\boldsymbol{\alpha}_x)$ is uniformly distributed over $[0, 1]$ [13, Thm. 15.2]. (Note the difference with respect to the phase-noise case where $\boldsymbol{\alpha}_x$ is uniformly distributed.) Intuitively, the pre-log is doubled because, in the absence of phase noise, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can be used to carry information.

D. Correlated Phase Noise at One Side or Diagonal Channel Matrix

Note that in the case of parallel phase-noise channels, we have an N-fold pre-log increase compared to the single-antenna case. In Section III-F, we shall show that this pre-log increase does not necessarily manifests itself for the general $N \times N$ MIMO phase-noise channel (2).

Before doing so, we first identify the conditions under which the pre-log of the channel in (2) coincides with the pre-log of the memoryless parallel phase-noise channel in (6), and, hence, MIMO multiplexing gain is present.

Proposition 3: The pre-log of the $N \times N$ MIMO phase-noise channel (2) is given by $N/2$ in the following three cases:
  i) both $\underline{\Phi}_T$ and $\underline{\Phi}_R$ have independent entries and the matrix $\underline{H}$ is a permutation of $\underline{I}_N$;
  ii) $\boldsymbol{\phi}_1^t = \cdots = \boldsymbol{\phi}_N^t$ (i.e., full correlation at the transmit side) and $\underline{\Phi}_R$ has independent entries;
  iii) $\boldsymbol{\phi}_1^r = \cdots = \boldsymbol{\phi}_N^r$ (i.e., full correlation at the receive side) and $\underline{\Phi}_T$ has independent entries.

Proof:

i) The matrix $\underline{\mathbf{H}}$ can be transformed into $\underline{\mathbf{I}}_N$ by a simple rearrangement of the order of the entries of $\underline{\mathbf{y}}$ and $\underline{\mathbf{x}}$. The proof is concluded by noting that $\underline{\boldsymbol{\Phi}}_R \underline{\boldsymbol{\Phi}}_T \sim \underline{\boldsymbol{\Phi}}$ (where $\underline{\boldsymbol{\Phi}}$ is defined as in Proposition 1) and by using Proposition 1.

ii) Let $\phi_1^t = \cdots = \phi_N^t = \phi$. Then

$$\underline{\mathbf{y}} = \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T \underline{\mathbf{x}} + \underline{\mathbf{w}}$$
$$= \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{H}} e^{j\phi} \underline{\mathbf{x}} + \underline{\mathbf{w}}$$
$$\sim \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{H}} \, \underline{\mathbf{x}} + \underline{\mathbf{w}}.$$

Since $\underline{\mathbf{H}}$ is unitary, if a probability distribution on $\underline{\mathbf{x}}$ satisfies (4), then the induced probability distribution on $\underline{\mathbf{H}} \underline{\mathbf{x}}$ satisfies (4) as well. Hence, using that $\underline{\mathbf{H}}$ is known to the transmitter, we have

$$\sup_{\mathsf{P}_{\underline{\mathbf{x}}}} I(\underline{\mathbf{x}}; \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{H}} \, \underline{\mathbf{x}} + \underline{\mathbf{w}}) = \sup_{\mathsf{P}_{\underline{\mathbf{x}}}} I(\underline{\mathbf{x}}; \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{x}} + \underline{\mathbf{w}}).$$

The proof then follows from Proposition 1.

iii) Let $\phi_1^r = \cdots = \phi_N^r = \phi$. Then

$$\underline{\mathbf{y}} = \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T \underline{\mathbf{x}} + \underline{\mathbf{w}}$$
$$= e^{j\phi} \underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T \underline{\mathbf{x}} + \underline{\mathbf{w}}$$
$$\sim \underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T \underline{\mathbf{x}} + \underline{\mathbf{w}}.$$

Since $\underline{\mathbf{H}}$ is unitary and known to the receiver, and since $\underline{\mathbf{H}}^H \underline{\mathbf{w}} \sim \underline{\mathbf{w}}$, we have

$$I(\underline{\mathbf{x}}; \underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T \underline{\mathbf{x}} + \underline{\mathbf{w}}) = I(\underline{\mathbf{x}}; \underline{\mathbf{H}}^H (\underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T \, \underline{\mathbf{x}} + \underline{\mathbf{w}}))$$
$$= I(\underline{\mathbf{x}}; \underline{\boldsymbol{\Phi}}_T \, \underline{\mathbf{x}} + \underline{\mathbf{w}}).$$

The proof then follows from Proposition 1. ∎

E. Correlated Phase Noise at Both Sides

We next consider the case where the phase-noise processes are perfectly correlated both at the transmitter and at the receiver side. In this case, the pre-log equals $N - 1/2$. Roughly speaking, $2N - 1$ out of the $2N$ real parameters characterizing $\underline{\mathbf{x}}$ can be used to carry information. This result is formalized in the following proposition.

Proposition 4: Consider the $N \times N$ MIMO phase-noise channel in (2) and assume that $\phi_1^t = \cdots = \phi_N^t = \phi^t$ and that $\phi_1^r = \cdots = \phi_N^r = \phi^r$. Then the capacity pre-log is given by $N - 1/2$.

Proof: Note that

$$\underline{\mathbf{y}} = \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T \underline{\mathbf{x}} + \underline{\mathbf{w}}$$
$$= e^{j\phi_r} e^{j\phi_t} \underline{\mathbf{H}} \underline{\mathbf{x}} + \underline{\mathbf{w}}$$
$$\sim e^{j\phi} \underline{\mathbf{H}} \, \underline{\mathbf{x}} + \underline{\mathbf{w}}.$$

where ϕ is uniformly distributed over $[0, 2\pi]$. Proceeding as in the proof of Proposition 3, we have that

$$\sup_{\mathsf{P}_{\underline{\mathbf{x}}}} I(\underline{\mathbf{y}}; \underline{\mathbf{x}}) = \sup_{\mathsf{P}_{\underline{\mathbf{x}}}} I(e^{j\phi} \underline{\mathbf{x}} + \underline{\mathbf{w}}; \underline{\mathbf{x}}). \tag{8}$$

The mutual information on the right-hand side (RHS) of (8) coincides with the mutual information of a block-memoryless phase-noise channel, with block-length equal to N. The desired result then follows from [14], [3]. ∎

F. Independent Phase Noise at Each Antenna

We next focus on the case of independent phase-noise processes across both transmit and receive antennas. Proposition 3, Case i) implies that MIMO multiplexing gains are present in this setup when $\underline{\mathbf{H}}$ is a permutation of the identity matrix. Since this situation is typically not encountered in LOS microwave backhaul links, in this section, we concentrate on channel matrices that do not satisfy this property. Throughout this section, we will deal exclusively with the 2×2 case. In fact, the proof of Theorem 5 below is tailored to the 2×2 case—an extension to the general $N \times N$ setup is currently under investigation. In view of Proposition 3, we shall assume that $\underline{\mathbf{H}}$ is not a permutation of $\underline{\mathbf{I}}_2$.

We start with two observations that follow directly from the system-model assumptions listed in Section II:

- After left-multiplying $\underline{\mathbf{x}}$ with $\underline{\boldsymbol{\Phi}}_T$, any information contained in the phases of the entries of $\underline{\mathbf{x}}$ is destroyed. Thus, the mutual information $I(\underline{\mathbf{x}}; \underline{\mathbf{y}})$ in (3) is not affected by the phase distribution of the entries of $\underline{\mathbf{x}}$. This property has already been used in the proof of Proposition 1.
- Symmetrically, because of left multiplication with $\underline{\boldsymbol{\Phi}}_R$, the phases of $\underline{\mathbf{y}}$ do not carry any information about $\underline{\mathbf{x}}$. Thus, the vector $|\underline{\mathbf{y}}|$ is a sufficient statistics for $\underline{\mathbf{x}}$. Hence, [12, Eq. (2.124)]

$$I(\underline{\mathbf{y}}; \underline{\mathbf{x}}) = I(|\underline{\mathbf{y}}|; |\underline{\mathbf{x}}|)$$

for every distribution $\mathsf{P}_{\underline{\mathbf{x}}}$.

While a characterization of the capacity pre-log for the 2×2 case is still out of reach, motivated by Lemma 2, we study here the rates achievable by input distributions under which $\|\underline{\mathbf{x}}\|^2$ and $\boldsymbol{\alpha}_x$ are independent and $\boldsymbol{\alpha}_x$ is uniformly distributed over $[0, \pi/2]$. We show that, in this case, the MIMO pre-log coincides with the single-antenna pre-log. Thus, in this case MIMO does not yield a pre-log increase compared to the single-antenna case.

Theorem 5: Consider the 2×2 MIMO phase-noise channel (2). Furthermore, assume that $\underline{\mathbf{H}}$ is not a permutation of $\underline{\mathbf{I}}_2$, and that both $\underline{\boldsymbol{\Phi}}_T$ and $\underline{\boldsymbol{\Phi}}_R$ have independent entries. Let $|\underline{\mathbf{x}}| = \|\underline{\mathbf{x}}\| \cdot [\cos(\boldsymbol{\alpha}_x) \, \sin(\boldsymbol{\alpha}_x)]^T$ with $\|\underline{\mathbf{x}}\|$ and $\boldsymbol{\alpha}_x$ independent, and with $\boldsymbol{\alpha}_x$ uniformly distributed over $[0, \pi/2]$. Then

$$\sup_{\mathsf{P}_{\|\underline{\mathbf{x}}\|}} I(|\underline{\mathbf{y}}|; |\underline{\mathbf{x}}|) \leq \frac{1}{2} \log(\rho) + \mathcal{O}(1), \quad \rho \to \infty.$$

Proof: Let $|\hat{\underline{\mathbf{x}}}| = [\cos(\boldsymbol{\alpha}_x) \, \sin(\boldsymbol{\alpha}_x)]^T$. Furthermore, define

$$\hat{\underline{\mathbf{r}}} \triangleq \underline{\boldsymbol{\Phi}}_R \underline{\mathbf{H}} \, \underline{\boldsymbol{\Phi}}_T |\hat{\underline{\mathbf{x}}}| \tag{9}$$

so that $\underline{\mathbf{y}} \sim \|\underline{\mathbf{x}}\| \hat{\underline{\mathbf{r}}} + \underline{\mathbf{w}}$. We start by noting that

$$I(\underline{\mathbf{y}}; |\underline{\mathbf{x}}|) = I(\underline{\mathbf{y}}; \|\underline{\mathbf{x}}\|, |\hat{\underline{\mathbf{x}}}|)$$
$$= I(\underline{\mathbf{y}}; \|\underline{\mathbf{x}}\|) + I(\underline{\mathbf{y}}; |\hat{\underline{\mathbf{x}}}| \mid \|\underline{\mathbf{x}}\|). \tag{10}$$

We next upper-bound the two terms on the RHS (10). For the first term, we have

$$
\begin{aligned}
I(\underline{\mathbf{y}}; \|\underline{\mathbf{x}}\|) &\leq I(\underline{\mathbf{y}}, \hat{\underline{\mathbf{r}}}; \|\underline{\mathbf{x}}\|) \\
&\overset{(a)}{=} I(\underline{\mathbf{y}}; \|\underline{\mathbf{x}}\| \,|\, \hat{\underline{\mathbf{r}}}) \\
&\overset{(b)}{=} I(\|\underline{\mathbf{x}}\| + \mathbf{n}; \|\underline{\mathbf{x}}\|) \\
&\overset{(c)}{\leq} \frac{1}{2}\log(1+\rho)
\end{aligned}
\tag{11}
$$

where (a) follows because $\hat{\underline{\mathbf{r}}}$ and $\|\underline{\mathbf{x}}\|$ are independent; (b) follows because $\|\underline{\mathbf{x}}\| + \mathbf{n}$ (with $\mathbf{n} \triangleq \Re\{\hat{\underline{\mathbf{r}}}^H \underline{\mathbf{w}}\}$) is a sufficient statistics for the detection of $\|\underline{\mathbf{x}}\|$ from $\underline{\mathbf{y}}$; and (c) follows because the Gaussian distribution maximizes differential entropy under a variance constraint.

For the second term on the RHS of (10), we have

$$
\begin{aligned}
I(\underline{\mathbf{y}}; |\hat{\underline{\mathbf{x}}}| \,|\, \|\underline{\mathbf{x}}\|) &\leq I(\underline{\mathbf{y}}, \underline{\mathbf{w}}; |\hat{\underline{\mathbf{x}}}| \,|\, \|\underline{\mathbf{x}}\|) \\
&\overset{(a)}{=} I(\underline{\mathbf{y}}; |\hat{\underline{\mathbf{x}}}| \,|\, \|\underline{\mathbf{x}}\|, \underline{\mathbf{w}}) \\
&= I(\hat{\underline{\mathbf{r}}}; |\hat{\underline{\mathbf{x}}}| \,|\, \|\underline{\mathbf{x}}\|, \underline{\mathbf{w}}) \\
&\overset{(b)}{=} I(\hat{\underline{\mathbf{r}}}; |\hat{\underline{\mathbf{x}}}|) \\
&\overset{(c)}{=} I(|\hat{\underline{\mathbf{r}}}|; |\hat{\underline{\mathbf{x}}}|)
\end{aligned}
$$

where (a) follows because $\underline{\mathbf{w}}$ and $\underline{\mathbf{x}}$ are independent; (b) follows because the pair $(\hat{\underline{\mathbf{r}}}, |\hat{\underline{\mathbf{x}}}|)$ is independent of $(\|\underline{\mathbf{x}}\|, \underline{\mathbf{w}})$; and (c) follows because the phase of the entries of $\hat{\underline{\mathbf{r}}}$ do not carry information about $|\hat{\underline{\mathbf{x}}}|$ [see (9)].

Note that $I(|\hat{\underline{\mathbf{r}}}|; |\hat{\underline{\mathbf{x}}}|)$ does not depend on $\mathsf{P}_{\|\underline{\mathbf{x}}\|}$, hence, it does not depend on ρ either. To conclude the proof it thus suffices to show that $I(|\hat{\underline{\mathbf{r}}}|; |\hat{\underline{\mathbf{x}}}|)$ is bounded. Let

$$
|\hat{\underline{\mathbf{r}}}| = \begin{bmatrix} \cos(\boldsymbol{\alpha}_r) \\ \sin(\boldsymbol{\alpha}_r) \end{bmatrix}
$$

for some $\boldsymbol{\alpha}_r \in [0, \pi/2]$. Furthermore, let h_{11} and h_{12} denote the first and the second entry, respectively, of the first row of $\underline{\mathrm{H}}$. It follows from (9) that

$$
\begin{aligned}
\cos^2(\boldsymbol{\alpha}_r) = &|h_{11}|^2 \cos^2(\boldsymbol{\alpha}_x) + |h_{12}|^2 \sin^2(\boldsymbol{\alpha}_x) \\
&+ 2\Big[|h_{11}|\,|h_{12}|\cos(\boldsymbol{\alpha}_x)\sin(\boldsymbol{\alpha}_x)\cos(\phi)\Big]
\end{aligned}
\tag{12}
$$

where ϕ is uniformly distributed over $[0, 2\pi]$. We further have that

$$
\begin{aligned}
I(|\hat{\underline{\mathbf{r}}}|; |\hat{\underline{\mathbf{x}}}|) &= I(\boldsymbol{\alpha}_r; \boldsymbol{\alpha}_x) \\
&= I(\cos^2(\boldsymbol{\alpha}_r); \boldsymbol{\alpha}_x)
\end{aligned}
$$

where, by the theorem's assumptions, $\boldsymbol{\alpha}_x$ is uniformly distributed over $[0, \pi/2]$. Let $\mathsf{q_u}(u)$ denote the uniform distribution over $[0, 1]$. It follows by *duality* (see Appendix) that

$$
\begin{aligned}
I(\cos^2(\boldsymbol{\alpha}_r); \boldsymbol{\alpha}_x) \leq &- \underbrace{\mathbb{E}_{\boldsymbol{\alpha}_r}\big[\log(\mathsf{q_u}(\cos^2(\boldsymbol{\alpha}_r)))\big]}_{=0} \\
&- h(\cos^2(\boldsymbol{\alpha}_r) \,|\, \boldsymbol{\alpha}_x) \\
= &- \mathbb{E}[\log(\sin(2\boldsymbol{\alpha}_x))] \\
&- \log(|h_{11}|\,|h_{12}|) - h(\cos(\phi))
\end{aligned}
\tag{13}
$$

where the last step follows from (12).

The proof is concluded by noting that, by the theorem's assumptions, $|h_{11}|$ and $|h_{12}|$ are strictly positive, so $\log(|h_{11}|\,|h_{12}|)$ is finite. Furthermore, $h(\cos(\phi))$ is finite because $\cos(\phi)$ follows an arcsine distribution. Finally,

$$
\begin{aligned}
\mathbb{E}[\log(\sin(2\boldsymbol{\alpha}_x))] &= \frac{2}{\pi}\int_0^{\pi/2} \log(\sin(2x))dx \\
&= -\log(2)
\end{aligned}
\tag{14}
$$

where the last equality follows from [15, Sec. 4.224, Eq. 3]. Theorem 5 follows by combining (10)–(14). $\blacksquare$

IV. CONCLUSIONS

We have demonstrated that MIMO does not yield a multiplexing gain when the phase noise is strong and independent across antennas, and when the direction of the input signal is isotropically distributed and independent of the magnitude. Although our result is not conclusive, it suggests that MIMO does not provide multiplexing gain in microwave backhaul links operating in the 20–40 GHz range. Does this mean that MIMO should not be used? Not necessarily. Even if MIMO does not yield an increase with respect to the pre-log (i.e., the first-order term in the high-SNR expansion of capacity), it could still give rise to an increase in the higher-order terms.[1] This might yield a noticeable throughput gain at SNR values of practical interest.

Our analysis was based on the assumptions that

i) the MIMO channel matrix $\underline{\mathrm{H}}$ is unitary and

ii) the phase-noise is i.i.d. and uniformly distributed.

It can be shown that the results reported in Propositions 3 and 4 continue to hold if Assumption i) is replaced by the weaker assumption that $\underline{\mathrm{H}}$ is full rank but not necessarily unitary (proof omitted for space constraint). This situation may arise in MIMO LOS systems when the antenna separation required for $\underline{\mathrm{H}}$ to be unitary is too large [11].

Propositions 3 and 4 continue to hold when Assumption ii) is replaced by the weaker assumption that the phase-noise process is stationary and has finite differential-entropy rate. The widely used Wiener phase-noise model [18] satisfies this assumption, provided that the initial phase distribution is uniform over $[0, 2\pi]$. Indeed, although memory in the phase-noise process increases capacity, this increase is bounded in SNR and does not affect the capacity pre-log. Specifically, let C_{Wiener} denote the capacity of a Wiener phase-noise channel and $C_{\mathrm{i.i.d.,\ unif}}$ denote the capacity of the memoryless uniform phase noise channel considered in this paper. Then [5], [19]

$$
C_{\mathrm{Wiener}}(\rho) = C_{\mathrm{i.i.d.,\ unif}}(\rho) + k_{\mathrm{Wiener}} + o(1), \quad \rho \to \infty.
\tag{15}
$$

Here, k_{Wiener} is a constant that depends on the memory in the phase-noise process, but does not depend on ρ. The capacity characterization (15) has been recently used to obtain nonasymptotic capacity bounds on the capacity of MIMO Wiener phase-noise channels [19].

---

[1] A result of this nature has been observed for the capacity of MIMO fading channels in the noncoherent setting [16], [17].

V. Acknowledgments

We wish to thank Mats Rydström, Thomas Ericsson, Sergio Benedetto, and Guido Montorsi for fruitful discussions.

Appendix

The proof of the inequality in (13) is based on *duality* [20], a technique that allows one to establish tight upper bounds on the mutual information between the input $\mathbf{x}$ of a channel and its output $\mathbf{y}$ by selecting an appropriate distribution on $\mathbf{y}$. Specifically, let $P_{\mathbf{x}}$ be a probability distribution on $\mathbf{x}$, let $P_{\mathbf{y}\,|\,\mathbf{x}}$ denote the conditional distribution on $\mathbf{y}$ given $\mathbf{x}$ (i.e., the channel law), and let $P_{\mathbf{y}}$ be the probability distribution induced on $\mathbf{y}$ by $P_{\mathbf{x}}$ through the channel law. Furthermore, let $Q_{\mathbf{y}}$ be an arbitrary probability distribution on $\mathbf{y}$ with pdf $q_{\mathbf{y}}$. We can upper-bound $I(\mathbf{x};\mathbf{y})$ using duality as follows [20, Thm. 5.1]:

$$
\begin{aligned}
I(\mathbf{x};\mathbf{y}) &= \mathbb{E}_{P_{\mathbf{x}}}\big[D(P_{\mathbf{y}\,|\,\mathbf{x}}\,\|\,P_{\mathbf{y}})\big] \\
&= \mathbb{E}_{P_{\mathbf{x}}}\big[D(P_{\mathbf{y}\,|\,\mathbf{x}}\,\|\,Q_{\mathbf{y}})\big] - D(P_{\mathbf{y}}\,\|\,Q_{\mathbf{y}}) \\
&\leq \mathbb{E}_{P_{\mathbf{x}}}\big[D(P_{\mathbf{y}\,|\,\mathbf{x}}\,\|\,Q_{\mathbf{y}})\big] \\
&= -\mathbb{E}_{P_{\mathbf{x}}}[\log(q_{\mathbf{y}}(\mathbf{y}))] - h(\mathbf{y}\,|\,\mathbf{x}). \qquad (16)
\end{aligned}
$$

Here, the first step follows by the definition of mutual information, the second step follows by a change of measure argument, the third step follows from the nonnegativity of the relative entropy, and the last step follows again by definition.

The main motivation for using the upper-bound (16) is that, for a given $P_{\mathbf{x}}$, the corresponding mutual information $I(\mathbf{x};\mathbf{y})$ might be difficult to compute. On the other hand, the RHS of (16) might be computable, provided that $Q_{\mathbf{y}}$ is chosen appropriately.

References

[1] F. Ivanek, "Mobile backhaul [from the guest editor's desk]," *IEEE Micro*, vol. 10, no. 5, pp. 10–20, Aug. 2009.

[2] 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.

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

[4] L. Barletta, M. Magarini, and A. Spalvieri, "Estimate of information rates of discrete-time first-order Markov phase noise channels," *IEEE Photon. Technol. Lett.*, vol. 23, no. 21, pp. 1582–1584, Nov. 2011.

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

[6] D. S. Baum and H. Bölcskei, "Information-theoretic analysis of MIMO channel sounding," *IEEE Trans. Inf. Theory*, pp. 7555–7577, Nov. 2011.

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

[8] Y. Zhu, L. Ni, and B. Belzer, "Design of turbo-coded modulation for the AWGN channel with Tikhonov phase error," vol. 53, no. 10, pp. 1697–1707, Oct. 2005.

[9] A. Barbieri, G. Colavolpe, and G. Caire, "Joint iterative detection and decoding in the presence of phase noise and frequency offset," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 171–179, Jan. 2007.

[10] A. Tarable, G. Montorsi, S. Benedetto, and S. Chinnici, "An EM-based phase-noise estimator for MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, to appear.

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

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

[13] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized algorithms for analysis and control of uncertain systems.* Springer, 2005.

[14] 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.

[15] I. Gradshteyn and I. Ryzhik, *Table of integrals, series and products*, 7th ed. Burlington, MA, U.S.A.: Elsevier, 2007.

[16] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 661–665.

[17] T. Koch, *On Heating Up and Fading in Communication Channels*, ser. Information Theory and its Applications, A. Lapidoth, Ed. Konstanz, Germany: Hartung-Gorre Verlag, May 2009, vol. 5.

[18] 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)*, Sep. 2012, pp. 316–327.

[19] 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)*, San Diego, CA, U.S.A., Feb. 2013.

[20] 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.