Information-Theoretic Analysis of MIMO Channel Sounding

Daniel S. Baum, *Member, IEEE,* and Helmut Bölcskei, *Fellow, IEEE*

***Abstract*—The large majority of commercially available multiple-input multiple-output (MIMO) radio channel measurement devices (sounders) is based on time-division multiplexed switching (TDMS) of a single transmit/receive radio frequency chain into the elements of a transmit/receive antenna array. While being cost-effective, such a solution can cause significant measurement errors due to phase noise and frequency offset in the local oscillators. In this paper, we systematically analyze the resulting errors and show that, in practice, *overestimation of channel capacity by several hundred percent* can occur. Overestimation is caused by phase noise (and to a lesser extent by frequency offset) leading to an increase of the MIMO channel rank. Our analysis furthermore reveals that the impact of phase errors is, in general, most pronounced if the physical channel has low rank (typical for line-of-sight or poor scattering scenarios). The extreme case of a rank-1 physical channel is analyzed in detail. The capacity bounds derived in this paper show excellent agreement with measurement results. In the light of the findings of this paper, the results obtained through MIMO channel measurement campaigns using TDMS-based channel sounders should be interpreted with great care.**

***Index Terms*—Channel measurement, sounding, multiple-input multiple-output (MIMO), phase noise.**

I. Introduction

THE USE of multiple-input multiple-output (MIMO) wireless communication promises significant improvements over existing wireless systems both in terms of spectral efficiency and link reliability. Obtaining accurate measurements of MIMO radio channels is of key importance to devising accurate MIMO radio channel models, which in turn are vital for system design, simulation, and performance analysis.

A common and widespread MIMO channel measurement device (a.k.a. sounder) design is based on time-division multiplexing with synchronous switching, or time-division multiplexed switching (TDMS) for short, of a single radio frequency (RF) chain into the individual elements of an antenna array. TDMS can be used at either the transmitter or the receiver (*one-sided TDMS*) or at both sides of the link (*double-sided TDMS*). For the latter case, which is practically the most relevant one, such an architecture is depicted in Fig. 1. TDMS leads to very cost-effective solutions as only a single RF chain is required at either the transmitter or the receiver (one-sided TDMS) or at both sides of the link (double-sided TDMS). To the best of our knowledge, the large majority of commercially available MIMO channel sounders is based on the TDMS principle. A major drawback of TDMS-based sounder architectures results from *temporal phase deviations* between the outputs of the local oscillators (LOs) in the RF chains at transmitter and receiver being *translated into the spatial domain* due to switching across antenna elements. This can cause an increase of the MIMO channel rank and corresponding measurement errors, in terms of estimated MIMO channel capacity, that can be on the order of several hundred percent. It is therefore immediately clear that understanding the impact of phase errors[1] in TDMS-based sounding is of fundamental importance.

One may argue that in a wireless communication link the impact of phase fluctuations in the transmitter and/or the receiver can simply be absorbed into an effective channel consisting of the physical propagation channel combined with LO-related (and potentially other) impairments. Channel estimation at the receiver for demodulation and decoding or for precoding at the transmitter (through feedback) would then simply work on the effective channel. This point of view can certainly be sensible in a data transmission setup if the frequency dispersion caused by phase fluctuations is small compared to that induced by the physical channel. In a channel sounding setup, however, it is crucial to separate the physical propagation channel (i.e., the object to be measured) from transmitter/receiver impairments, in order to obtain measurement results that depend as little as possible on the measurement device (sounder) used. Furthermore, as already pointed out, the measurement procedure employed in TDMS-based MIMO channel sounding results in very high sensitivity of the estimated channel capacity with respect to (w.r.t.) phase errors.

Contributions: The goal of this paper is to systematically analyze the impact of phase noise and frequency offset (between transmitter and receiver LO) on estimated MIMO channel capacity when TDMS-based channel sounders are used. In particular, we show that the presence of phase errors can lead to significant overestimation of MIMO channel capacity. A sensitivity analysis reveals that, in certain cases, underestimation is possible as well, albeit typically resulting in significantly smaller errors.

Our analysis is based on the signal model devised in [1]. This model is applicable to the wide class of correlation-

Manuscript received September ?, 2007; revised December 3, 2002. The paper was presented in part at IEEE VTC Fall 2004, Sept. 2004, Los Angeles, CA. This work was supported in part by the Swiss National Science Foundation (SNF) under grant No. 200021-100025/1.

D. S. Baum was with ETH Zurich, 8092 Zurich, Switzerland. He is now with Baum Consulting GmbH, 8003 Zurich, Switzerland (e-mail: daniel.baum@ieee.org).

H. Bölcskei is with ETH Zurich, 8092 Zurich, Switzerland (e-mail: boelcskei@nari.ee.ethz.ch).

Communicated by [TBD]

Digital Object Identifier [TBD]

[1]For brevity, in the remainder of this paper, we use the terminology *phase errors* whenever we refer to phase deviations due to phase noise, or frequency offset, or both.

based (as defined in [2, Sec. III]) MIMO channel sounders and takes into account phase noise and frequency offset. We then systematically identify situations where phase errors have no (or little) impact on MIMO (ergodic and outage) channel capacity estimates and where they lead to significant estimation errors. As an extreme case in the latter category, we demonstrate that even moderate phase noise can turn a rank-1 physical channel (e.g., a pin-hole channel [3], [4]) into a full-rank effective channel.

The capacity bounds derived in this paper are found to exhibit an excellent match with measurement results, specifically those reported in the companion paper [1].

Previous related work: For fully parallel MIMO channel sounders, i.e., channel sounders employing a separate RF chain for each transmit and each receive antenna element, the impact of gain imbalance in parallel RF chains and of thermal noise on the estimated capacity of a physical rank-1 MIMO channel is analyzed in [5]. The variance of an approximation of the error in the mutual information (MI) of an effective channel resulting from a deterministic MIMO channel subject to additive white complex Gaussian distributed perturbations is derived in [6]. For physical pin-hole [3], a.k.a. key-hole [4], [7], [8] (i.e., rank-1), MIMO channels in a controlled indoor environment, the impact of measurement imperfections such as thermal noise and "multi-path leakage" (i.e., multi-path components propagating between the transmitter and the receiver via paths other than through the pin-hole) on the channel eigenvalue distribution and the resulting outage capacity are analyzed numerically, based on measurements and simulations, in [8].

Organization of the paper: The remainder of this paper is organized as follows. In Section II, the architecture of a TDMS-based MIMO channel sounder is briefly reviewed and the corresponding channel and signal model are introduced. In Section III, we study the effect of phase errors on estimated MI. A framework for analyzing the sensitivity of the MIMO channel MI to phase errors is developed in Section IV. Finally, Section V is devoted to the special (but practically relevant) case of rank-1 physical channels. We conclude in Section VI.

Notation: $\mathbb{E}\{\cdot\}$ denotes the expectation operator. The Dirac delta function is denoted as $\delta_i = 1$ for $i = 0$ and 0 otherwise. The superscripts ^{T}, ^{H}, and * stand for transposition, conjugate transposition, and elementwise conjugation, respectively. For a complex scalar $z \in \mathbb{C}$, the functions $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ stand for the real and imaginary part of z, respectively.

An $m \times n$ matrix is a matrix with m rows and n columns. $\mathbf{1}$ and $\mathbf{0}$ denote an all-ones and all-zeros matrix, respectively, of appropriate dimensions. If required, the dimensions of a matrix are specified through subscripts, e.g., $\mathbf{1}_{m,n}$. $\mathbf{I}_m$ stands for the $m \times m$ identity matrix. $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard (pointwise) product of the matrices $\mathbf{A}$ and $\mathbf{B}$, and $f^{\circ}(\mathbf{A})$ stands for the matrix resulting from entry-wise application of the function $f(\cdot)$ to $\mathbf{A}$. $\|\mathbf{A}\|_{\mathrm{F}}$ is the Frobenius norm of $\mathbf{A}$ and $\mathrm{Tr}(\mathbf{A})$ is the trace of $\mathbf{A}$. $\mathrm{diag}(\mathbf{x})$ denotes the diagonal matrix with the elements of the vector $\mathbf{x}$ on its main diagonal, and $\mathrm{dg}(\mathbf{A}) = \mathbf{A} \circ \mathbf{I}$ zeros out all but the diagonal elements of $\mathbf{A}$. The element in the mth row and nth column of $\mathbf{A}$ is denoted as $[\mathbf{A}]_{m,n}$. $\mathrm{r}(\mathbf{A})$ and $\lambda_i(\mathbf{A})$ stand for the rank and the ith eigenvalue of $\mathbf{A}$, respectively. Unless explicitly stated otherwise, eigenvalues are sorted in decreasing order, i.e., $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$. For an $m \times n$ matrix $\mathbf{A} = [\,\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n\,]$ with columns $\mathbf{a}_i$, we define the $mn \times 1$ vector $\mathrm{vec}(\mathbf{A}) = [\,\mathbf{a}_1^T \;\; \mathbf{a}_2^T \;\; \cdots \;\; \mathbf{a}_n^T\,]^T$. The commutation matrix $\mathbf{K}_{(m,n)}$ [9, Sec. 3.7] is a permutation matrix of size $mn \times mn$ uniquely defined through

$$\mathbf{K}_{(m,n)}\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\mathbf{A}^T) \tag{1}$$

where $\mathbf{A}$ is an $m \times n$ matrix. For brevity, we define $\mathrm{divec}(\mathbf{A}) = \mathrm{diag}\big(\mathrm{vec}(\mathbf{A})\big)$.

For two random variables (RVs) X and Y, $X \overset{\mathrm{d}}{=} Y$ and $X \overset{\mathrm{d}}{\approx} Y$ stands for equivalence and approximate equivalence in distribution, respectively. For a RV X, the probability density function (pdf) and cumulative distribution function (cdf) are denoted by $\mathrm{p}_X(x)$ and $\mathrm{F}_X(x)$, respectively. The variance of a RV X is denoted as $\mathrm{Var}\{X\}$. The covariance matrix of a complex random matrix $\mathbf{X}$ is defined as $\mathrm{Cov}\{\mathbf{X}\} = \mathbb{E}\big\{\big(\mathrm{vec}(\mathbf{X}) - \mathbb{E}\{\mathrm{vec}(\mathbf{X})\}\big)\big(\mathrm{vec}(\mathbf{X}) - \mathbb{E}\{\mathrm{vec}(\mathbf{X})\}\big)^H\big\}$.

A real Gaussian random vector is defined as a vector with jointly Gaussian (JG) elements; the corresponding distribution is denoted by $\mathcal{N}(\mathbf{m}, \mathbf{C})$, where $\mathbf{m}$ is the mean and $\mathbf{C}$ is the covariance matrix. A complex Gaussian random vector is defined as a vector with JG real and imaginary parts. A complex random vector will be called proper if its pseudo-covariance matrix vanishes. $\mathcal{CN}(\mathbf{m}, \mathbf{C})$ stands for a proper complex Gaussian random vector with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$. The complex random vectors $\mathbf{x}$ and $\mathbf{y}$ will be called jointly proper if the composite random vector having $\mathbf{x}$ and $\mathbf{y}$ as subvectors is proper.

For a chi-square distributed RV with n degrees of freedom and variance $2n\sigma^4$ we write χ^2_{n,σ^2}, where $\chi^2_{n,\sigma^2} \overset{\mathrm{d}}{=} \|\mathbf{x}\|^2$ with $\mathbf{x} \overset{\mathrm{d}}{=} \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$. $[\,x_1 \;\; x_2 \;\; \cdots \;\; x_n\,] \overset{\mathrm{d}}{=} D_n(a_1, a_2, \ldots, a_n)$, $x_i \geq 0$ $(i = 1, 2, \ldots, n)$ denotes a Dirichlet distributed random vector with parameters $a_1, a_2, \ldots, a_n$. The corresponding subvector $\mathbf{x} = [\,x_1 \;\; x_2 \;\; \cdots \;\; x_{n-1}\,]$ satisfies $\mathbf{x} \overset{\mathrm{d}}{=} D_{n-1}(a_1, a_2, \ldots, a_{n-1}; a_n)$ and has joint pdf [10, Th. 1.2]

$$\mathrm{p}_{D_{n-1}(a_1,a_2,\ldots,a_{n-1};a_n)}(\mathbf{x}) = \frac{\Gamma(a)}{\prod_{i=1}^{n}\Gamma(a_i)}\left(\prod_{i=1}^{n-1} x_i^{a_i - 1}\right)\left(1 - \sum_{i=1}^{n-1} x_i\right)^{a_n - 1}$$

where $a = \sum_{i=1}^{n} a_i$ and $\Gamma(z)$ is the Gamma function [11, Sec. 6.1].

We denote the beta distribution with parameters a and b as $\beta(a, b)$ with pdf given by

$$\mathrm{p}_{\beta(a,b)}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(1-x)^{b-1}x^{a-1}, \quad a > 0,\ b > 0,\ x \geq 0.$$

The digamma function $\Psi(z)$ is defined as [11, Eq. 6.3.1, Eq. 6.3.16]

$$\begin{aligned} \Psi(z) &= \frac{\mathrm{d}}{\mathrm{d}z}\log_e\Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \psi_0(z), \quad z \in \mathbb{C} \\ \Psi(z) &= -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n+z-1)}, \quad z \neq 0, -1, -2, \ldots \end{aligned} \tag{2}$$

where $\gamma \approx 0.5772$ is Euler's constant, and $\psi_n(z)$ is the

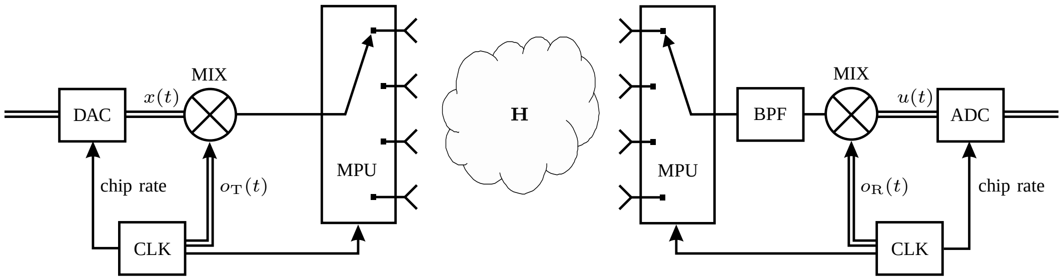

Fig. 1. Architecture of a TDMS-based MIMO channel sounder. ADC, DAC, MIX, MPU, BPF, and CLK stand for analog-to-digital converter, digital-to-analog converter, mixer, multiplexing unit, bandpass filter, and (reference) clock, respectively.

polygamma function [11, Eq. 6.4.1], defined as the nth derivative of $\Psi(z)$. We will also need the following representation of the digamma function at positive integer multiples of $1/2$ given by [11, Eq. 6.3.2, Eq. 6.3.4]

$$\begin{aligned} \Psi(k) &= -\gamma + \sum_{n=1}^{k-1} \frac{1}{n}, \\ \Psi\left(k - \frac{1}{2}\right) &= -\gamma - 2\log_e(2) + \sum_{n=1}^{k-1} \frac{2}{2n-1}, \\ & \qquad\qquad k \in \mathbb{N}, k \geq 1. \end{aligned} \tag{3}$$

In addition, the first derivative of the digamma function can be written as an infinite series as [11, Eq. 6.4.10]

$$\begin{aligned} \Psi'(z) = \psi_1(z) &= \frac{\mathrm{d}^2}{\mathrm{d}z^2} \log_e \Gamma(z) \\ &= \sum_{n=1}^{\infty} \frac{1}{(n+z-1)^2}, \\ & \qquad\qquad z \neq 0, -1, -2, \ldots. \end{aligned} \tag{4}$$

Finally, we note that all logarithms are to the base 2 unless explicitly stated otherwise. Throughout the paper, the number of transmit and receive antenna elements in a MIMO channel is denoted as M_T and M_R, respectively, and we will refer to antenna elements simply as antennas.

II. MIMO Channel Sounding Based On Time-Division Multiplexed Switching

In this section, we describe the system architecture of a correlation-based MIMO channel sounder employing TDMS and we present the corresponding signal model, taking into account the presence of phase errors.

A. Channel Sounder Architecture

The basic architecture of a TDMS-based MIMO channel sounder is depicted in Fig. 1. A (possibly complex) sounding signal $x(t)$ is generated in baseband and modulated to the propagation channel center frequency. At the receiver, bandpass filtering (to remove out-of-band thermal noise) and downconversion to baseband, resulting in the signal $u(t)$, is followed by extraction of the MIMO channel estimate. Both the transmitter and the receiver employ a multiplexing unit, which steps a single RF chain through all transmit/receive antennas sequentially in time following a prescribed switching pattern. Clocks at transmitter and receiver serve as reference for the analog-to-digital converter (ADC) and digital-to-analog converter (DAC) sampling rates, antenna multiplexing timing, and RF mixing (i.e., LO) frequencies.

Another frequently used MIMO channel sounder setup employs a single antenna (either at the transmitter or the receiver or at both sides of the link), which is physically moved (in an automated fashion) to form a "virtual" antenna array. This setup also fits into the framework described in the paper. The impact of phase errors on MIMO capacity estimates in such a time-division-based "virtual" antenna array sounder architecture will, in general, be significantly more pronounced than in the TDMS-based case where a single RF chain is switched electronically into different physical antenna elements. This is because in the "virtual" antenna array case, the time that passes when moving the single antenna from a given physical position to the next one is much longer than the time it takes to switch electronically between different antenna elements.

B. SISO Signal Model

We start by briefly describing the single-input single-output (SISO) signal model which constitutes the basis for the MIMO signal model in TDMS-based sounders introduced in Section II-C. A detailed discussion of the signal model for correlation-based sounders can be found in [1].

Throughout the paper, we shall work in complex baseband. The SISO snapshot (measurement) time instants are given by t_k, $k = 1, 2, \ldots$, and are referred to as SISO snapshot times in the following. The quantities $t_k - t_{k-1}$ are referred to as the SISO snapshot time distances. We assume that the (time-varying) physical channel to be measured is linear time-invariant (LTI) during each SISO snapshot interval (the time interval during which the channel is measured to obtain a SISO snapshot; by definition less than the SISO snapshot time distance) and can change from SISO snapshot interval to SISO snapshot interval.

For flat-fading physical channels, considered in the remainder of the paper, we simply get

$$\hat{h}_k = h_k \, e^{j\mu_k} \, e^{j\varphi_k} \tag{5}$$

where we set $\mu_k = \Delta\omega t_k$ with $\Delta\omega = 2\pi(f_\mathrm{T} - f_\mathrm{R})$ denoting the frequency offset between the independent clock sources at transmitter and receiver. The model (5) is standard and well-known in the literature.

C. MIMO Signal Model

The basic principle of TDMS-based MIMO channel sounding is to sequentially measure the $M_\mathrm{T}M_\mathrm{R}$ scalar subchannels of the MIMO channel. Since the individual SISO subchannels (note that we continue to use the term *SISO snapshot* to refer to the measurement of one subchannel of the MIMO matrix) are band-limited stochastic processes (due to finite Doppler spread), it is not necessary to assume that the subchannels are static during the entire MIMO measurement period. Rather, it suffices to choose the sampling rate of the MIMO channel (which is also the rate at which SISO snapshots are taken for each SISO subchannel) in compliance with the sampling theorem and to properly align the measurements in time [12]. In the remainder of the paper, we assume that this alignment has already been performed.

Denoting the *effective* (i.e, the physical channel including the effect of phase errors) scalar subchannel between the nth ($n = 1, 2, \ldots, M_\mathrm{T}$) transmit and the mth ($m = 1, 2, \ldots, M_\mathrm{R}$) receive antenna as $\hat{h}_{m,n} = h_{m,n} \exp(j(\mu_{m,n} + \varphi_{m,n}))$, the corresponding effective MIMO channel matrix can be expressed as

$$\widehat{\mathbf{H}} = \mathbf{H} \circ \underbrace{\exp^\circ\big(j(\mathbf{M} + \mathbf{\Phi})\big)}_{\mathbf{\Theta}} \tag{6}$$

where $[\mathbf{M}]_{m,n} = \mu_{m,n}$ and $[\mathbf{\Phi}]_{m,n} = \varphi_{m,n}$. What we would like to measure is the *physical* channel matrix $\mathbf{H}$. However, due to phase errors, the sounder has access to the effective channel matrix $\widehat{\mathbf{H}}$ only. The entries in $\mathbf{M}$ and $\mathbf{\Phi}$ depend on the switching pattern, i.e., the order in which the individual scalar subchannels are measured, and the SISO snapshot time distances. In the following, we denote a switching pattern as an ordered sequence of pairs $\big((m_1, n_1), (m_2, n_2), \ldots, (m_{M_\mathrm{T}M_\mathrm{R}}, n_{M_\mathrm{T}M_\mathrm{R}})\big)$ where (m_k, n_k) means that the scalar subchannel h_{m_k,n_k} is being measured at time t_k.

We conclude this section by introducing an approximation that will frequently be used. For small phase errors, we use the standard first-order Taylor-series approximation (see, e.g., [13, Eq. (4.12)])

$$\exp^\circ(j\mathbf{\Phi}) \approx \mathbf{1} + j\mathbf{\Phi}. \tag{7}$$

In the remainder of the paper, whenever referring to $\widehat{\mathbf{H}}$, unless explicitly stated otherwise, we shall use the exact expression for $\mathbf{\Theta}$ according to (6). We conclude this section by noting that, throughout the paper, whenever we deal with random physical channels, $\mathbf{\Theta}$ will be assumed to be statistically independent of $\mathbf{H}$.

III. Effect of Phase Errors on Mutual Information

The purpose of this section is to analyze the impact of phase errors on MI for random physical channels. Analytic results for the general (arbitrary rank (with probability (w.p.) 1) of the physical channel) case seem very difficult to obtain. We therefore start by identifying cases where the MI is not affected even though the channel statistics are. Analytic results for deterministic physical channels and for (deterministic or random) rank-1 physical channels will be provided in Sections IV and V, respectively.

We analyze a MIMO channel with input-output relation

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

where $\mathbf{s}$ is the $M_\mathrm{T} \times 1$ transmit vector, $\mathbf{r}$ is the $M_\mathrm{R} \times 1$ receive vector, and $\mathbf{n}$ is an $M_\mathrm{R} \times 1$ noise vector distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{M_\mathrm{R}})$. We assume no channel state information (CSI) at the transmitter,[2] perfect CSI at the receiver, and take the input signal vector to be circularly-symmetric complex Gaussian with covariance matrix $\mathbf{Q} = (\rho/M_\mathrm{T})\mathbf{I}_{M_\mathrm{T}}$, where ρ is the average signal-to-noise ratio (SNR) at each of the receive antennas. The MI (in bit/s/Hz) of this channel is therefore given by [14]

$$\begin{aligned} I &= \log\det\left(\mathbf{I}_{M_\mathrm{R}} + \frac{\rho}{M_\mathrm{T}}\mathbf{H}\mathbf{H}^H\right) \\ &= \log\det\left(\mathbf{I}_{M_\mathrm{T}} + \frac{\rho}{M_\mathrm{T}}\mathbf{H}^H\mathbf{H}\right). \end{aligned} \tag{8}$$

The purpose of this section is to study how I changes when $\mathbf{H}$ in (8) is replaced by $\widehat{\mathbf{H}}$ in (6), i.e., to analyze the statistics of $\hat{I} = \log\det\big(\mathbf{I}_{M_\mathrm{R}} + (\rho/M_\mathrm{T})\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H\big) = \log\det\big(\mathbf{I}_{M_\mathrm{T}} + (\rho/M_\mathrm{T})\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\big)$. The statistics of MI are the basis for various quantities that characterize limits of transmission rates as briefly discussed in the following.

The (Shannon) capacity in the ergodic setting and the outage capacity in the non-ergodic setting are obtained by optimizing the covariance matrix of the input signal vector, subject to the transmit power constraint.

For the ergodic setting, the optimal input covariance matrix for the general (w.r.t. the channel statistics) case has not been characterized so far. Interesting first results in this direction appear in [15], where the capacity-achieving input distribution for the Unitary-Independent-Unitary (UIU) MIMO channel model is found. Unfortunately, the UIU channel model does not encompass the MIMO channel model in the presence of phase noise as used here. The characterization of the optimal input covariance matrix in this case appears to be rather difficult.

For the outage setting, even less is known in terms of optimum input covariances, i.e., input covariance matrices, that lead to minimum outage probability. Even for the classical independent and identically distributed (i.i.d.) Rayleigh-fading MIMO channel model, the problem of finding the outage-capacity minimizing covariance is open. Telatar's conjecture [14] states that the transmit power is to be distributed equally among a subset of the transmit streams and that the number of "active modes" depends on the SNR.

In this paper, we therefore assume that $\mathbf{Q} = (\rho/M_\mathrm{T})\mathbf{I}_{M_\mathrm{T}}$. The corresponding mutual information (8) leads to a lower-bound on ergodic capacity and outage capacity, respectively, which we call ergodic and outage *system capacity*, respectively. For i.i.d. Rayleigh-fading channels, the ergodic system capacity coincides with Shannon capacity.

The system capacity has practical relevance for the following reasons. To achieve Shannon capacity, the transmitter needs to determine the optimum transmit signal covariance matrix, derived from the channel statistics. Providing the transmitter with perfect knowledge of the channel statistics can be difficult

[2]While the transmitter is not assumed to know the channel realization, it still needs to know the channel statistics.

in practice. Even if this knowledge is available, implementing a transmitter that adjusts the transmit signal statistics to the propagation environment is non-trivial. Moreover, to come close to Shannon capacity in practice, signal shaping is needed to make the discrete transmit signal alphabet look as Gaussian as possible. Signal shaping becomes more involved in the MIMO case, as shaping needs to ensure that the transmit vector looks *jointly* Gaussian with the desired covariance matrix. It is therefore not surprising that setting $\mathbf{Q} = (\rho/M_\mathrm{T})\mathbf{I}_{M_\mathrm{T}}$, commonly referred to as spatial multiplexing, results in a transmission method that is widely used in practice.

For the high-SNR case, the optimum transmit signal covariance matrix needs to satisfy $\mathrm{r}(\mathbf{Q}) \geq \mathrm{r}(\mathbf{H})$ as $\log\det\big(\mathbf{I}_{M_\mathrm{R}} + \mathbf{H}\mathbf{Q}\mathbf{H}^H\big) = \mathrm{r}(\mathbf{H}\mathbf{Q}\mathbf{H}^H)\log(\rho/M_\mathrm{T}) + O(1)$, and $\mathrm{r}(\mathbf{H}\mathbf{Q}\mathbf{H}^H) = \mathrm{r}(\mathbf{H}^H\mathbf{H}\mathbf{Q}) \leq \min(\mathrm{r}(\mathbf{H}), \mathrm{r}(\mathbf{Q}))$. Choosing $\mathbf{Q} = (\rho/M_\mathrm{T})\mathbf{I}_{M_\mathrm{T}}$ therefore yields the high-SNR Shannon (and outage) capacity up to an $O(1)$-term. As we are often interested in the channel rank increase caused by phase noise, it follows immediately that system capacity provides us with the full picture.

In the remainder of the paper, for simplicity of exposition, we simply use the terms (ergodic) capacity and outage capacity when referring to ergodic and outage system capacity, respectively.

A. Cases Where Mutual Information is Not Affected

We start by showing that in the special case of an i.i.d. purely Rayleigh fading physical channel $\mathbf{H}$, i.e., $\mathbf{H}$ being a zero-mean proper (or equivalently "circularly symmetric") complex Gaussian random matrix, the channel is not affected by phase errors in the sense that the effective channel $\widehat{\mathbf{H}}$ is i.i.d. Rayleigh fading as well, irrespectively of the statistics of the phase errors.

Defining $\mathbf{h} = \mathrm{vec}(\mathbf{H}) \overset{\mathrm{d}}{=} \mathcal{CN}(\mathbf{0}, \mathbf{I})$, we want to show that $\mathbf{D}\mathbf{h} \overset{\mathrm{d}}{=} \mathcal{CN}(\mathbf{0}, \mathbf{I})$, where $\mathbf{D} = \mathrm{diag}\big(\big[\exp\big(j(\mu_1 + \varphi_1)\big) \;\; \exp\big(j(\mu_2 + \varphi_2)\big) \;\; \cdots \;\; \exp\big(j(\mu_{M_\mathrm{T}M_\mathrm{R}} + \varphi_{M_\mathrm{T}M_\mathrm{R}})\big)\big]\big)$. We start by noting that the characteristic function of $\mathbf{h}$ is given by [16, Eq. (20)]

$$\begin{aligned}\Psi_\mathbf{h}(j\boldsymbol{\nu}) &= \mathbb{E}\big\{e^{j\mathrm{Re}(\mathbf{h}^H\boldsymbol{\nu})}\big\} \\ &= e^{-\boldsymbol{\nu}^H\boldsymbol{\nu}/4}, \quad \boldsymbol{\nu} \in \mathbb{C}^{M_\mathrm{T}M_\mathrm{R}}.\end{aligned}$$

The characteristic function of $\mathbf{D}\mathbf{h}$ is obtained as

$$\begin{aligned}\Psi_{\mathbf{D}\mathbf{h}}(j\boldsymbol{\nu}) &= \mathbb{E}\big\{e^{j\mathrm{Re}(\mathbf{h}^H\mathbf{D}^H\boldsymbol{\nu})}\big\} = \mathbb{E}_\mathbf{D}\big\{\mathbb{E}_{\mathbf{h}|\mathbf{D}}\big\{e^{j\mathrm{Re}(\mathbf{h}^H\mathbf{D}^H\boldsymbol{\nu})}\big\}\big\} \\ &= \mathbb{E}_\mathbf{D}\big\{e^{-\boldsymbol{\nu}^H\mathbf{D}\mathbf{D}^H\boldsymbol{\nu}/4}\big\} = \mathbb{E}_\mathbf{D}\big\{e^{-\boldsymbol{\nu}^H\boldsymbol{\nu}/4}\big\} \\ &= e^{-\boldsymbol{\nu}^H\boldsymbol{\nu}/4} = \Psi_\mathbf{h}(j\boldsymbol{\nu})\end{aligned}$$

where we made use of the fact that $\mathbf{D}\mathbf{D}^H = \mathbf{I}$. We have therefore shown that i.i.d. purely Rayleigh fading physical channels are not affected by phase errors.

Furthermore, for correlated Rayleigh or Ricean fading physical channels (the latter being represented by a non-zero-mean proper complex Gaussian random matrix), phase errors can have a significant impact on the channel statistics and thus on the MI. Corresponding numerical results are provided in [1]. There are cases, however, where even though the statistics of the effective channel $\widehat{\mathbf{H}}$ differ from the statistics of $\mathbf{H}$, we still have $\hat{I} = I$. Intuitively, this happens because of the quadratic dependence of $\hat{I}$ on $\widehat{\mathbf{H}}$.

1) The low-SNR regime: Phase errors have no impact on MI in the low-SNR regime, irrespectively of the physical channel's statistics and the phase error statistics. This can easily be seen by noting that for low SNR

$$I \approx \log\bigg(1 + \frac{\rho}{M_\mathrm{T}}\|\mathbf{H}\|_\mathrm{F}^2\bigg)$$

which, combined with the fact that $\|\mathbf{H}\|_\mathrm{F}^2 = \|\widehat{\mathbf{H}}\|_\mathrm{F}^2$, proves the statement. In general, we can conclude that the impact of phase errors on MI is more pronounced for higher SNR. This is because phase errors lead to a rank increase of the MIMO channel and high-SNR MI depends strongly on the rank (or multiplexing gain) of the channel, whereas low-SNR MI depends only on the Frobenius norm $\|\mathbf{H}\|_\mathrm{F}$.

2) One-sided switching or fully parallel sounding: For MIMO channel sounders where either the transmitter or the receiver employs one RF chain per antenna (i.e., parallel sounding) and hence no switching is necessary on the corresponding side of the link, the effective channel matrix is given by $\widehat{\mathbf{H}} = \mathbf{D}_\mathrm{R}\mathbf{H}d_\mathrm{T}$ and $\widehat{\mathbf{H}} = d_\mathrm{R}\mathbf{H}\mathbf{D}_\mathrm{T}$ in the case of switching only at the receive and the transmit side, respectively. Here, d_R and d_T as well as the entries of the diagonal matrices $\mathbf{D}_\mathrm{R}$ and $\mathbf{D}_\mathrm{T}$ are of the form $\exp\big(j(\mu + \varphi)\big)$. Even though the statistics of the effective channel $\widehat{\mathbf{H}}$ are clearly different from the statistics of the physical channel $\mathbf{H}$, it is easily seen by direct insertion into (8) that the MI is not affected by "one-sided" phase errors. Obviously, this is also true for fully parallel (both the transmitter and the receiver employ one RF chain per antenna) MIMO channel sounders.

Throughout this paper, we employ a *regular sounding pattern* (where sounding pattern denotes the combination of a switching pattern and a set of SISO snapshot times)

$$\begin{aligned}m_k &= (k-1) \bmod M_\mathrm{R} + 1 \\ n_k &= (k-1) \operatorname{div} M_\mathrm{R} + 1 \\ t_k &= T_\mathrm{R}(m_k - 1) + T_\mathrm{T}(n_k - 1)\end{aligned} \tag{9}$$

with $k = 1, 2, \ldots, M_\mathrm{T}M_\mathrm{R}$, characterized by the timing parameters $T_\mathrm{T}, T_\mathrm{R} \in \mathbb{R}^+$. This corresponds to starting with transmit antenna 1, switching through the receive antennas $1, 2, \ldots, M_\mathrm{R}$ sequentially, then switching to transmit antenna 2, again switching through the receive antennas sequentially with the same SISO snapshot time distances between the receive antennas as before, and so on. Equivalently, we could start with receive antenna 1, switch through the transmit antennas sequentially and so on. The impact of using other switching patterns is discussed in [1].

B. Example

To illustrate the effect of phase errors as well as different switching patterns, consider a simple example with $M_\mathrm{T} = M_\mathrm{R} = 2$. The switching pattern $\big((1,1), (2,1), (2,2), (1,2)\big)$ leads to

$$\boldsymbol{\Theta}_1 = \begin{bmatrix} e^{j(\mu_1+\varphi_1)} & e^{j(\mu_4+\varphi_4)} \\ e^{j(\mu_2+\varphi_2)} & e^{j(\mu_3+\varphi_3)} \end{bmatrix}$$

whereas the switching pattern $\big((1,1),(2,1),(1,2),(2,2)\big)$ results in

$$\boldsymbol{\Theta}_2 = \begin{bmatrix} e^{j(\mu_1+\varphi_1)} & e^{j(\mu_3+\varphi_3)} \\ e^{j(\mu_2+\varphi_2)} & e^{j(\mu_4+\varphi_4)} \end{bmatrix}.$$

The physical channel matrix $\mathbf{H}$ is, of course, unaffected by the switching pattern. The dependence of $\boldsymbol{\Theta}$ on the switching pattern and on the SISO snapshot time distances (which through frequency offset correspond to phase differences) is highly problematic since different switching patterns and/or SISO snapshot times yield different (incorrect) measurement results for the same physical MIMO channel. Assume that the physical channel is given by $\mathbf{H} = \mathbf{1}$, and $\mu_1 + \varphi_1 = \mu_4 + \varphi_4 = 0$, $\mu_2 + \varphi_2 = -\pi/2$, and $\mu_3 + \varphi_3 = \pi/2$. It is then easily seen that $\lambda_1\big((\mathbf{H}\circ\boldsymbol{\Theta}_1)(\mathbf{H}\circ\boldsymbol{\Theta}_1)^H\big) = \lambda_2\big((\mathbf{H}\circ\boldsymbol{\Theta}_1)(\mathbf{H}\circ\boldsymbol{\Theta}_1)^H\big) = 2$ whereas $\lambda_1\big((\mathbf{H}\circ\boldsymbol{\Theta}_2)(\mathbf{H}\circ\boldsymbol{\Theta}_2)^H\big) = 4$ and $\lambda_2\big((\mathbf{H}\circ\boldsymbol{\Theta}_2)(\mathbf{H}\circ\boldsymbol{\Theta}_2)^H\big) = 0$. In summary, starting from a rank-1 physical channel, depending on the switching pattern, we can get a rank-1 or a rank-2 effective channel.

IV. Sensitivity Analysis

The purpose of this section is twofold. First, we introduce a tool for evaluating the sensitivity of the MI of a fixed physical channel to phase errors. This will be accomplished by computing the first two terms in the Taylor series expansion of $\hat{I}(\mathbf{X})$ with $\mathbf{X} = \mathbf{M} + \boldsymbol{\Phi}$ around the phase-error-free case $\mathbf{X} = \mathbf{0}$. Based on this framework, we will then be able to provide analytic expressions for approximations of the first and second moment of the MI of the effective channel for arbitrary phase noise covariance matrix. For the sake of simplicity of exposition, we assume that $M_\mathrm{R} \le M_\mathrm{T}$.

We shall be concerned with computing the second-order Taylor series expansion of

$$\hat{I}(\mathbf{X}) = \log\det\left(\mathbf{I}_{M_\mathrm{R}} + \frac{\rho}{M_\mathrm{T}}\big(\mathbf{H}\circ\exp^\circ(j\mathbf{X})\big)\big(\mathbf{H}\circ\exp^\circ(j\mathbf{X})\big)^H\right)$$

around $\mathbf{X} = \mathbf{0}$ given by

$$\tilde{I}(\mathbf{X}) = \hat{I}(\mathbf{0}) + \mathcal{J}_{\hat{I}}(\mathbf{0})\,\mathrm{vec}(\mathbf{X}) + \frac{1}{2}\,\big(\mathrm{vec}(\mathbf{X})\big)^T \mathcal{H}_{\hat{I}}(\mathbf{0})\,\mathrm{vec}(\mathbf{X}) \tag{10}$$

where $\mathcal{J}_{\hat{I}}(\mathbf{0})$ denotes the $1\times M_\mathrm{T}M_\mathrm{R}$ Jacobian matrix (vector, in this case) and $\mathcal{H}_{\hat{I}}(\mathbf{0})$ is the $M_\mathrm{T}M_\mathrm{R}\times M_\mathrm{T}M_\mathrm{R}$ Hessian matrix of $\hat{I} = \hat{I}(\mathbf{X})$ at $\mathbf{X} = \mathbf{0}$. Clearly, we have $\hat{I}(\mathbf{0}) = I$. Even though the second-order Taylor series expansion of MI does not yield accurate approximations in the case of rank-deficient physical channels, analytic expressions for the Jacobian matrix $\mathcal{J}_{\hat{I}}(\mathbf{0})$ and the Hessian matrix $\mathcal{H}_{\hat{I}}(\mathbf{0})$ can be used to test whether $\hat{I}(\mathbf{X})$ has an extremum at $\mathbf{X} = \mathbf{0}$.

Even though the computation of the Jacobian matrix $\mathcal{J}_{\hat{I}}(\mathbf{0})$ and the Hessian matrix $\mathcal{H}_{\hat{I}}(\mathbf{0})$ does not pose any major technical difficulties, it still requires the application of tools that are not completely standard, namely matrix differential calculus [9] and matrix-variate Wirtinger a.k.a. $\mathbb{CR}$ calculus as described in [17]. We shall therefore present the corresponding derivations in some detail.

Theorem 1: The $1\times M_\mathrm{T}M_\mathrm{R}$ Jacobian matrix (vector) $\mathcal{J}_{\hat{I}}(\mathbf{0})$ in (10) is given by

$$\mathcal{J}_{\hat{I}}(\mathbf{0}) = 2\log(e)\,\mathrm{Im}\bigg(\Big(\mathrm{vec}\big((\mathbf{Y}^{-1}\mathbf{H})\circ\mathbf{H}^*\big)\Big)^T\bigg) \tag{11}$$

where

$$\mathbf{Y} = \frac{M_\mathrm{T}}{\rho}\mathbf{I}_{M_\mathrm{R}} + \mathbf{H}\mathbf{H}^H.$$

The $M_\mathrm{T}M_\mathrm{R} \times M_\mathrm{T}M_\mathrm{R}$ Hessian matrix $\mathcal{H}_{\hat{I}}(\mathbf{0})$ in (10) is given by

$$\begin{aligned}\mathcal{H}_{\hat{I}}(\mathbf{0}) = 2\log(e)&\\ \cdot\,\mathrm{Re}\Big(&\mathrm{divec}(\mathbf{H})\mathbf{K}_{(M_\mathrm{T},M_\mathrm{R})}\big((\mathbf{H}^H\mathbf{Y}^{-1})^T\otimes\mathbf{H}^H\mathbf{Y}^{-1}\big)\\ &\qquad\cdot\,\mathrm{divec}(\mathbf{H})\\ &+\mathrm{divec}(\mathbf{H})\big((\mathbf{I}_{M_\mathrm{T}} - \mathbf{H}^H\mathbf{Y}^{-1}\mathbf{H})\otimes(\mathbf{Y}^{-1})^T\big)\\ &\qquad\cdot\,\mathrm{divec}(\mathbf{H}^*)\\ &-\mathrm{divec}\big((\mathbf{Y}^{-1}\mathbf{H})\circ\mathbf{H}^*\big)\Big).\end{aligned} \tag{12}$$

Proof: We start by defining

$$\widehat{\mathbf{Y}} = \frac{M_\mathrm{T}}{\rho}\mathbf{I}_{M_\mathrm{R}} + \underbrace{\big(\mathbf{H}\circ\exp^\circ(j\mathbf{X})\big)}_{\widehat{\mathbf{H}}}\underbrace{\big(\mathbf{H}\circ\exp^\circ(j\mathbf{X})\big)^H}_{\widehat{\mathbf{H}}^H}$$

so that $\mathbf{Y} = \widehat{\mathbf{Y}}|_{\mathbf{X}=\mathbf{0}}$ and hence $\hat{I} = \log\det\big((\rho/M_\mathrm{T})\widehat{\mathbf{Y}}\big)$. The strategy used in the proof of both parts of the statement is to compute $\mathrm{d}\hat{I}$ and $\mathrm{d}^2\hat{I}$ and to bring the resulting expressions into the form

$$\begin{aligned}\mathrm{d}\hat{I} &= \mathbf{A}\,\mathrm{vec}(\mathrm{d}\mathbf{X})\\ \mathrm{d}^2\hat{I} &= \big(\mathrm{vec}(\mathrm{d}\mathbf{X})\big)^T\mathbf{B}\,\mathrm{vec}(\mathrm{d}\mathbf{X})\end{aligned} \tag{13}$$

which will then allow us to apply the first [9, Ch. 5, Th. 6] and the second [9, Ch. 6, Th. 6] identification theorem for a real-valued function of real-valued parameters to conclude that

$$\mathcal{J}_{\hat{I}}(\mathbf{0}) = \mathbf{A} \qquad \text{and} \qquad \mathcal{H}_{\hat{I}}(\mathbf{0}) = \frac{1}{2}(\mathbf{B} + \mathbf{B}^T).$$

Computing the Jacobian matrix: Using the basic rules of differentiation together with the relation $\mathrm{d}\log_e\det(\mathbf{A}) = \mathrm{Tr}(\mathbf{A}^{-1}\,\mathrm{d}\mathbf{A})$ [9, Sec. 8.3, Eq. (11)], we obtain

$$\mathrm{d}\log\det\left(\frac{\rho}{M_\mathrm{T}}\widehat{\mathbf{Y}}\right) = \log(e)\,\mathrm{Tr}\Big(\widehat{\mathbf{Y}}^{-1}\,\mathrm{d}\widehat{\mathbf{Y}}\Big).$$

Applying the product rule $\mathrm{d}(\mathbf{A}\mathbf{B}) = (\mathrm{d}\mathbf{A})\mathbf{B} + \mathbf{A}\mathrm{d}\mathbf{B}$ [9, Sec. 8.2, Eq. (15)] and $\mathrm{d}(\mathbf{A}^T) = (\mathrm{d}\mathbf{A})^T$ [9, Sec. 8.2, Eq. (18)], we get, using Wirtinger calculus,

$$\mathrm{d}\widehat{\mathbf{Y}} = \mathrm{d}(\widehat{\mathbf{H}}\,\widehat{\mathbf{H}}^H) = (\mathrm{d}\widehat{\mathbf{H}})\,\widehat{\mathbf{H}}^H + \widehat{\mathbf{H}}\,\big(\mathrm{d}(\widehat{\mathbf{H}}^*)\big)^T.$$

With $\mathrm{d}(\mathbf{A}\circ\mathbf{B}) = (\mathrm{d}\mathbf{A})\circ\mathbf{B} + \mathbf{A}\circ(\mathrm{d}\mathbf{B})$ [9, Sec. 8.2, Eq. (17)], we have $\mathrm{d}\widehat{\mathbf{H}} = \mathrm{d}(\mathbf{H}\circ\boldsymbol{\Theta}) = \mathbf{H}\circ\mathrm{d}\boldsymbol{\Theta}$ and $\mathrm{d}(\widehat{\mathbf{H}}^*) = \mathbf{H}^*\circ\mathrm{d}(\boldsymbol{\Theta}^*)$. Noting that $[\mathrm{d}\big(f^\circ(\mathbf{X})\big)]_{m,n} = \mathrm{d}\big([f^\circ(\mathbf{X})]_{m,n}\big) = \mathrm{d}\big(f([\mathbf{X}]_{m,n})\big)$, we obtain

$$\mathrm{d}\log\det\left(\frac{\rho}{M_\mathrm{T}}\widehat{\mathbf{Y}}\right) = \log(e)\,\mathrm{Tr}\bigg(\widehat{\mathbf{Y}}^{-1}\left(\frac{1}{j}\widehat{\mathbf{H}}\,\big(\widehat{\mathbf{H}}\circ\mathrm{d}\mathbf{X}\big)^H - \frac{1}{j}\big(\widehat{\mathbf{H}}\circ\mathrm{d}\mathbf{X}\big)\,\widehat{\mathbf{H}}^H\right)\bigg)$$

and further

$$\begin{aligned}
&\mathrm{d}\log\det\left(\frac{\rho}{M_{\mathrm{T}}}\widehat{\mathbf{Y}}\right)\\
&\quad=\log(e)\operatorname{Tr}\Bigg(\widehat{\mathbf{Y}}^{-1}\left(\frac{1}{j}\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}\circ\mathrm{d}\mathbf{X}\right)^{H}\right.\\
&\qquad\qquad\left.-\frac{1}{j}\left(\widehat{\mathbf{H}}^{*}\left(\widehat{\mathbf{H}}\circ\mathrm{d}\mathbf{X}\right)^{H}\right)^{*}\right)\Bigg)\\
&\quad=2\log(e)\operatorname{Im}\left(\operatorname{Tr}\left(\widehat{\mathbf{Y}}^{-1}\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}\circ\mathrm{d}\mathbf{X}\right)^{H}\right)\right)\\
&\quad=2\log(e)\operatorname{Im}\left(\operatorname{Tr}\left((\widehat{\mathbf{Y}}^{-1}\widehat{\mathbf{H}})^{T}\left(\widehat{\mathbf{H}}^{*}\circ\mathrm{d}\mathbf{X}\right)\right)\right). \qquad (14)
\end{aligned}$$

It remains to turn (14) into the form $\mathrm{d}\hat{I}=\mathbf{A}\operatorname{vec}(\mathrm{d}\mathbf{X})$. This can be done by first showing that

$$\operatorname{Tr}\big(\mathbf{A}(\mathbf{B}\circ\mathbf{C})\big)=\big(\operatorname{vec}(\mathbf{A}^{T}\circ\mathbf{B})\big)^{T}\operatorname{vec}(\mathbf{C}) \qquad (15)$$

and then applying (15) to (14). In order to prove (15), we start by noting that with $\operatorname{Tr}(\mathbf{A}^{T}\mathbf{B})=\operatorname{vec}(\mathbf{A})^{T}\operatorname{vec}(\mathbf{B})$ [9, Sec. 2.4, Eq. (4)], we have

$$\operatorname{Tr}\big(\mathbf{A}(\mathbf{B}\circ\mathbf{C})\big)=\big(\operatorname{vec}(\mathbf{A}^{T})\big)^{T}\operatorname{vec}(\mathbf{B}\circ\mathbf{C})$$

which upon application of

$$\operatorname{vec}(\mathbf{A}\circ\mathbf{B})=\operatorname{divec}(\mathbf{A})\operatorname{vec}(\mathbf{B}) \qquad (16)$$

yields the desired result. Finally, applying (15) to (14), we obtain

$$\begin{aligned}
&\mathrm{d}\log\det\left(\frac{\rho}{M_{\mathrm{T}}}\widehat{\mathbf{Y}}\right)\\
&\quad=2\log(e)\operatorname{Im}\left(\left(\operatorname{vec}\big((\mathbf{Y}^{-1}\mathbf{H})\circ\mathbf{H}^{*}\big)\right)^{T}\right)\operatorname{vec}(\mathrm{d}\mathbf{X})
\end{aligned}$$

which proves (11).

Computing the Hessian matrix: We start by noting that

$$\begin{aligned}
&\mathrm{d}^{2}\log\det\left(\frac{\rho}{M_{\mathrm{T}}}\widehat{\mathbf{Y}}\right)\\
&\quad=\log(e)\ \mathrm{d}\operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\,\mathrm{d}\widehat{\mathbf{Y}})\\
&\quad=\log(e)\ \operatorname{Tr}\big(\widehat{\mathbf{Y}}^{-1}\,(\mathrm{d}^{2}\widehat{\mathbf{Y}})-(\widehat{\mathbf{Y}}^{-1}\,\mathrm{d}\widehat{\mathbf{Y}})^{2}\big) \qquad (17)
\end{aligned}$$

where we used $\mathrm{d}\operatorname{Tr}(\mathbf{A})=\operatorname{Tr}(\mathrm{d}\mathbf{A})$ [9, Sec. 8.2, Eq. (20)] and $\mathrm{d}(\mathbf{A}^{-1})=-\mathbf{A}^{-1}(\mathrm{d}\mathbf{A})\mathbf{A}^{-1}$ [9, Sec. 8.4, Eq. (1)] along with the product rule. In order to keep the following exposition simple, we set $\mathrm{d}\widehat{\mathbf{H}}=j\widehat{\mathbf{H}}\circ\mathrm{d}\mathbf{X}=j\dot{\mathbf{H}}$ so that $\mathrm{d}\widehat{\mathbf{Y}}=\mathrm{d}(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^{H})=j\dot{\mathbf{H}}\,\widehat{\mathbf{H}}^{H}-j\widehat{\mathbf{H}}\,\dot{\mathbf{H}}^{H}$. Expanding the second term in (17) through similar manipulations as in the derivation of the Jacobian matrix, and using $(\widehat{\mathbf{Y}}^{-1})^{T}=(\widehat{\mathbf{Y}}^{-1})^{*}$, we get

$$\begin{aligned}
&-\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1}\mathrm{d}\widehat{\mathbf{Y}})^{2}\big)\\
&\quad=\operatorname{Tr}\big(\widehat{\mathbf{Y}}^{-1}(\dot{\mathbf{H}}\,\widehat{\mathbf{H}}^{H}-\widehat{\mathbf{H}}\,\dot{\mathbf{H}}^{H})\widehat{\mathbf{Y}}^{-1}(\dot{\mathbf{H}}\,\widehat{\mathbf{H}}^{H}-\widehat{\mathbf{H}}\,\dot{\mathbf{H}}^{H})\big)\\
&\quad=\operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\dot{\mathbf{H}}\,\widehat{\mathbf{H}}^{H}\widehat{\mathbf{Y}}^{-1}\dot{\mathbf{H}}\,\widehat{\mathbf{H}}^{H})\\
&\qquad-\operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\dot{\mathbf{H}}\,\widehat{\mathbf{H}}^{H}\widehat{\mathbf{Y}}^{-1}\widehat{\mathbf{H}}\,\dot{\mathbf{H}}^{H})\\
&\qquad+\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1})^{*}\dot{\mathbf{H}}^{*}\,\widehat{\mathbf{H}}^{T}\,(\widehat{\mathbf{Y}}^{-1})^{*}\dot{\mathbf{H}}^{*}\,\widehat{\mathbf{H}}^{T}\big)\\
&\qquad-\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1})^{*}\dot{\mathbf{H}}^{*}\,\widehat{\mathbf{H}}^{T}\,(\widehat{\mathbf{Y}}^{-1})^{*}\widehat{\mathbf{H}}^{*}\,\dot{\mathbf{H}}^{T}\big)\\
&\quad=2\operatorname{Re}\left(\operatorname{Tr}(\widehat{\mathbf{H}}^{H}\widehat{\mathbf{Y}}^{-1}\dot{\mathbf{H}}\widehat{\mathbf{H}}^{H}\,\widehat{\mathbf{Y}}^{-1}\dot{\mathbf{H}})\right)\\
&\qquad-2\operatorname{Re}\left(\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1})^{T}\dot{\mathbf{H}}^{*}\widehat{\mathbf{H}}^{T}\,(\widehat{\mathbf{Y}}^{-1})^{*}\widehat{\mathbf{H}}^{*}\dot{\mathbf{H}}^{T}\big)\right). \qquad (18)
\end{aligned}$$

Next, applying [9, Ch. 2, Th. 3]

$$\begin{aligned}
\operatorname{Tr}(\mathbf{ABCD})&=\big(\operatorname{vec}(\mathbf{D}^{T})\big)^{T}(\mathbf{C}^{T}\otimes\mathbf{A})\operatorname{vec}(\mathbf{B})\\
&=\big(\operatorname{vec}(\mathbf{D})\big)^{T}(\mathbf{A}\otimes\mathbf{C}^{T})\operatorname{vec}(\mathbf{B}^{T})
\end{aligned} \qquad (19)$$

with $\mathbf{A}=\widehat{\mathbf{H}}^{H}\widehat{\mathbf{Y}}^{-1}$, $\mathbf{B}=\dot{\mathbf{H}}$, $\mathbf{C}=\widehat{\mathbf{H}}^{H}\,\widehat{\mathbf{Y}}^{-1}$, and $\mathbf{D}=\dot{\mathbf{H}}$ to the first term in (18) and with $\mathbf{A}=(\widehat{\mathbf{Y}}^{-1})^{T}$, $\mathbf{B}=\dot{\mathbf{H}}^{*}$, $\mathbf{C}=\widehat{\mathbf{H}}^{T}\,(\widehat{\mathbf{Y}}^{-1})^{*}\widehat{\mathbf{H}}^{*}$, and $\mathbf{D}=\dot{\mathbf{H}}^{T}$ to the second term in (18), we obtain

$$\begin{aligned}
&-\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1}\mathrm{d}\widehat{\mathbf{Y}})^{2}\big)\\
&\quad=2\operatorname{Re}\Big(\big(\operatorname{vec}(\dot{\mathbf{H}}^{T})\big)^{T}\widehat{\mathbf{R}}_{1}\operatorname{vec}(\dot{\mathbf{H}})\Big)\\
&\qquad\qquad-2\operatorname{Re}\Big(\big(\operatorname{vec}(\dot{\mathbf{H}}^{T})\big)^{T}\widehat{\mathbf{R}}_{2}\operatorname{vec}(\dot{\mathbf{H}}^{H})\Big)\\
&\quad\overset{\text{(a)}}{=}2\left(\operatorname{vec}\big((\mathrm{d}\mathbf{X})^{T}\big)\right)^{T}\operatorname{Re}\big(\operatorname{divec}(\widehat{\mathbf{H}}^{T})\,\widehat{\mathbf{R}}_{1}\\
&\qquad\qquad\cdot\operatorname{divec}(\widehat{\mathbf{H}})\big)\operatorname{vec}(\mathrm{d}\mathbf{X})\\
&\qquad-2\left(\operatorname{vec}\big((\mathrm{d}\mathbf{X})^{T}\big)\right)^{T}\operatorname{Re}\big(\operatorname{divec}(\widehat{\mathbf{H}}^{T})\,\widehat{\mathbf{R}}_{2}\\
&\qquad\qquad\cdot\operatorname{divec}(\widehat{\mathbf{H}}^{H})\big)\operatorname{vec}\big((\mathrm{d}\mathbf{X})^{T}\big)
\end{aligned} \qquad (20)$$

where we set

$$\begin{aligned}
\widehat{\mathbf{R}}_{1}&=(\widehat{\mathbf{H}}^{H}\,\widehat{\mathbf{Y}}^{-1})^{T}\otimes\widehat{\mathbf{H}}^{H}\,\widehat{\mathbf{Y}}^{-1}\\
&=(\widehat{\mathbf{Y}}^{-1}\otimes\widehat{\mathbf{H}}^{*})^{T}(\widehat{\mathbf{H}}^{*}\otimes\widehat{\mathbf{Y}}^{-1})
\end{aligned}$$

and

$$\widehat{\mathbf{R}}_{2}=(\widehat{\mathbf{Y}}^{-1})^{T}\otimes\widehat{\mathbf{H}}^{H}\,\widehat{\mathbf{Y}}^{-1}\,\widehat{\mathbf{H}},$$

and we used $\dot{\mathbf{H}}=\widehat{\mathbf{H}}\circ\mathrm{d}\mathbf{X}$ and (16) in (a). Next, we need to rewrite (20) in terms of $\big(\operatorname{vec}(\mathrm{d}\mathbf{X})\big)^{T}$ and $\operatorname{vec}(\mathrm{d}\mathbf{X})$ only, which requires getting rid of the terms $(\mathrm{d}\mathbf{X})^{T}$ inside the $\big(\operatorname{vec}(\cdot)\big)^{T}$.

Upon applying (1) in (20), we obtain

$$\begin{aligned}
&-\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1}\mathrm{d}\widehat{\mathbf{Y}})^{2}\big)\\
&\quad=2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^{T}\mathbf{K}^{T}_{(M_{\mathrm{R}},M_{\mathrm{T}})}\\
&\qquad\qquad\cdot\operatorname{Re}\big(\operatorname{divec}(\widehat{\mathbf{H}}^{T})\,\widehat{\mathbf{R}}_{1}\operatorname{divec}(\widehat{\mathbf{H}})\big)\operatorname{vec}(\mathrm{d}\mathbf{X})\\
&\qquad-2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^{T}\mathbf{K}^{T}_{(M_{\mathrm{R}},M_{\mathrm{T}})}\\
&\qquad\qquad\cdot\operatorname{Re}\big(\operatorname{divec}(\widehat{\mathbf{H}}^{T})\,\widehat{\mathbf{R}}_{2}\operatorname{divec}(\widehat{\mathbf{H}}^{H})\big)\\
&\qquad\qquad\qquad\cdot\mathbf{K}_{(M_{\mathrm{R}},M_{\mathrm{T}})}\operatorname{vec}(\mathrm{d}\mathbf{X})
\end{aligned}$$

which using $\mathbf{K}^{T}_{(m,n)}=\mathbf{K}_{(n,m)}$ [9, Sec. 3.7, Eq. (2)] and

$$\begin{aligned}
&\big(\operatorname{vec}(\mathbf{A})\big)^{T}\mathbf{K}_{(n,m)}\operatorname{divec}(\mathbf{B})\\
&\quad=\big(\operatorname{vec}(\mathbf{A})\big)^{T}\operatorname{divec}(\mathbf{B}^{T})\,\mathbf{K}_{(n,m)}
\end{aligned} \qquad (21)$$

results in

$$\begin{aligned}
&-\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1}\mathrm{d}\widehat{\mathbf{Y}})^{2}\big)\\
&\quad=2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^{T}\operatorname{Re}\big(\operatorname{divec}(\widehat{\mathbf{H}})\mathbf{K}_{(M_{\mathrm{T}},M_{\mathrm{R}})}\\
&\qquad\qquad\cdot\widehat{\mathbf{R}}_{1}\operatorname{divec}(\widehat{\mathbf{H}})\big)\operatorname{vec}(\mathrm{d}\mathbf{X})\\
&\qquad-2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^{T}\operatorname{Re}\big(\operatorname{divec}(\widehat{\mathbf{H}})\mathbf{K}_{(M_{\mathrm{T}},M_{\mathrm{R}})}\\
&\qquad\qquad\cdot\widehat{\mathbf{R}}_{2}\,\mathbf{K}_{(M_{\mathrm{R}},M_{\mathrm{T}})}\operatorname{divec}(\widehat{\mathbf{H}}^{*})\big)\operatorname{vec}(\mathrm{d}\mathbf{X}).
\end{aligned} \qquad (22)$$

The validity of (21) can easily be seen by noting that $\operatorname{divec}(\mathbf{B})\,\mathbf{K}_{(m,n)}\operatorname{vec}(\mathbf{A})=\operatorname{divec}(\mathbf{B})\operatorname{vec}(\mathbf{A}^{T})=\operatorname{vec}(\mathbf{B}\,\circ$

$\mathbf{A}^T) = \operatorname{vec}\big((\mathbf{B}^T \circ \mathbf{A})^T\big) = \mathbf{K}_{(m,n)} \operatorname{vec}(\mathbf{B}^T \circ \mathbf{A}) = \mathbf{K}_{(m,n)} \operatorname{divec}(\mathbf{B}^T) \operatorname{vec}(\mathbf{A})$.

Finally, employing [9, Sec. 3.7, Eq. (5)]

$$\mathbf{K}_{(p,m)}(\mathbf{A} \otimes \mathbf{B})\mathbf{K}_{(n,q)} = \mathbf{B} \otimes \mathbf{A} \tag{23}$$

to the $m \times n$ matrix $\mathbf{A}$ and the $p \times q$ matrix $\mathbf{B}$, we can simplify the second term on the right-hand side (RHS) of (22) to obtain

$$\begin{aligned} &-\operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1}\mathrm{d}\widehat{\mathbf{Y}})^2\big) = \\ &\quad 2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^T \operatorname{Re}\Big(\operatorname{divec}(\widehat{\mathbf{H}})\mathbf{K}_{(M_\mathrm{T},M_\mathrm{R})} \\ &\quad\quad \cdot \big((\widehat{\mathbf{H}}^H\,\widehat{\mathbf{Y}}^{-1})^T \otimes (\widehat{\mathbf{H}}^H\,\widehat{\mathbf{Y}}^{-1})\big) \operatorname{divec}(\widehat{\mathbf{H}}) \\ &\quad - \operatorname{divec}(\widehat{\mathbf{H}})\big((\widehat{\mathbf{H}}^H\,\widehat{\mathbf{Y}}^{-1}\,\widehat{\mathbf{H}}) \otimes (\widehat{\mathbf{Y}}^{-1})^T\big) \\ &\quad\quad \cdot \operatorname{divec}(\widehat{\mathbf{H}}^*)\Big)\operatorname{vec}(\mathrm{d}\mathbf{X}). \end{aligned} \tag{24}$$

It remains to turn the first term in (17) into the form of the RHS of (13). To this end, we start by noting that, using Wirtinger calculus,

$$\begin{aligned} \mathrm{d}^2\widehat{\mathbf{Y}} &= \mathrm{d}\,\mathrm{d}(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H) = \mathrm{d}\big(j\dot{\mathbf{H}}\,\widehat{\mathbf{H}}^H - j\widehat{\mathbf{H}}\,\dot{\mathbf{H}}^H\big) \\ &= -\ddot{\mathbf{H}}\,\widehat{\mathbf{H}}^H + \dot{\mathbf{H}}\,\dot{\mathbf{H}}^H + \dot{\mathbf{H}}\,\dot{\mathbf{H}}^H - \widehat{\mathbf{H}}\,\ddot{\mathbf{H}}^H \end{aligned} \tag{25}$$

where we set $\mathrm{d}^2\widehat{\mathbf{H}} = -\widehat{\mathbf{H}} \circ \mathrm{d}\mathbf{X} \circ \mathrm{d}\mathbf{X} = -\ddot{\mathbf{H}}$. Next, inserting (25) into the first term of (17), we get

$$\operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\,\mathrm{d}^2\widehat{\mathbf{Y}}) \tag{26}$$

$$\begin{aligned} &= 2\operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\,\dot{\mathbf{H}}\,\dot{\mathbf{H}}^H) - \operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\,\ddot{\mathbf{H}}\,\widehat{\mathbf{H}}^H) \\ &\qquad - \operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\,\widehat{\mathbf{H}}\,\ddot{\mathbf{H}}^H) \\ &= 2\operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\,\dot{\mathbf{H}}\,\dot{\mathbf{H}}^H) - \operatorname{Tr}(\widehat{\mathbf{Y}}^{-1}\,\ddot{\mathbf{H}}\,\widehat{\mathbf{H}}^H) \\ &\qquad - \operatorname{Tr}\big((\widehat{\mathbf{Y}}^{-1})^*\,\ddot{\mathbf{H}}^*\,\widehat{\mathbf{H}}^T\big) \end{aligned}$$

$$= 2\operatorname{Tr}\Big(\widehat{\mathbf{Y}}^{-1}\dot{\mathbf{H}}\dot{\mathbf{H}}^H - \operatorname{Re}\big(\widehat{\mathbf{Y}}^{-1}\ddot{\mathbf{H}}\widehat{\mathbf{H}}^H\big)\Big) \tag{27}$$

$$\begin{aligned} &= 2\operatorname{Tr}\big(\widehat{\mathbf{Y}}^{-1}\dot{\mathbf{H}}\mathbf{I}_{M_\mathrm{T}}\dot{\mathbf{H}}^H\big) - 2\operatorname{Re}\Big(\operatorname{Tr}\big(\widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1}\ddot{\mathbf{H}}\big)\Big) \\ &\overset{\text{(a)}}{=} 2\left(\operatorname{vec}(\dot{\mathbf{H}}^T)\right)^T\big((\widehat{\mathbf{Y}}^{-1})^T \otimes \mathbf{I}_{M_\mathrm{T}}\big)\operatorname{vec}(\dot{\mathbf{H}}^H) \\ &\qquad - 2\operatorname{Re}\bigg(\Big(\operatorname{vec}\big((\widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1})^T\big)\Big)^T\operatorname{vec}(\ddot{\mathbf{H}})\bigg) \\ &\overset{\text{(b)}}{=} 2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^T\mathbf{K}_{(M_\mathrm{T},M_\mathrm{R})}\operatorname{divec}(\widehat{\mathbf{H}}^T)\big((\widehat{\mathbf{Y}}^{-1})^T \otimes \mathbf{I}_{M_\mathrm{T}}\big) \\ &\qquad \cdot \operatorname{divec}(\widehat{\mathbf{H}}^H)\,\mathbf{K}_{(M_\mathrm{R},M_\mathrm{T})}\operatorname{vec}(\mathrm{d}\mathbf{X}) \\ &\quad - 2\operatorname{Re}\bigg(\Big(\operatorname{vec}\big((\widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1})^T\big)\Big)^T \\ &\qquad \cdot \operatorname{divec}(\widehat{\mathbf{H}})\operatorname{divec}(\mathrm{d}\mathbf{X})\operatorname{vec}(\mathrm{d}\mathbf{X})\bigg) \end{aligned}$$

$$\begin{aligned} &\overset{\text{(c)}}{=} 2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^T\operatorname{divec}(\widehat{\mathbf{H}})\left(\mathbf{I}_{M_\mathrm{T}} \otimes (\widehat{\mathbf{Y}}^{-1})^T\right) \\ &\qquad \cdot \operatorname{divec}(\widehat{\mathbf{H}}^*)\operatorname{vec}(\mathrm{d}\mathbf{X}) \\ &\quad - 2\left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^T\operatorname{Re}\Big(\operatorname{divec}\big((\widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1})^T \circ \widehat{\mathbf{H}}\big)\Big) \\ &\qquad \cdot \operatorname{vec}(\mathrm{d}\mathbf{X}) \end{aligned} \tag{28}$$

where (a) results from applying (19) with $\mathbf{A} = \widehat{\mathbf{Y}}^{-1}$, $\mathbf{B} = \dot{\mathbf{H}}$, $\mathbf{C} = \mathbf{I}_{M_\mathrm{T}}$, and $\mathbf{D} = \dot{\mathbf{H}}^H$ to the first term, transposing the result, and, as before, applying $\operatorname{Tr}(\mathbf{A}^T\mathbf{B}) = \operatorname{vec}(\mathbf{A})^T\operatorname{vec}(\mathbf{B})$ with $\mathbf{A}^T = \widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1}$ and $\mathbf{B} = \ddot{\mathbf{H}}$ to the second term. Step (b) is a consequence of applying (16), the commutation relation (1), and $\mathbf{K}_{(m,n)}^T = \mathbf{K}_{(n,m)}$. To obtain (c), we used (21) and (23) for the first term and

$$\begin{aligned} &\left(\operatorname{vec}\big((\widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1})^T\big)\right)^T\operatorname{divec}(\widehat{\mathbf{H}})\operatorname{divec}(\mathrm{d}\mathbf{X}) \\ &\quad = \left(\operatorname{vec}\big((\widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1})^T\big)\right)^T\operatorname{divec}(\mathrm{d}\mathbf{X})\operatorname{divec}(\widehat{\mathbf{H}}) \\ &\quad = \left(\operatorname{vec}(\mathrm{d}\mathbf{X})\right)^T\operatorname{divec}\big((\widehat{\mathbf{H}}^H\widehat{\mathbf{Y}}^{-1})^T \circ \widehat{\mathbf{H}}\big) \end{aligned}$$

for the second term. The final result follows by identifying (28) and (24) with the RHS of (13), noting that $\widehat{\mathbf{H}} = \mathbf{H}$ for $\mathbf{X} = \mathbf{0}$, and applying the second identification theorem. All the terms, except for the first term in (28), can be verified to be real-symmetric[3] so that $(1/2)(\mathbf{B} + \mathbf{B}^T) = \mathbf{B}$. The first term in (28) is Hermitian and hence $(1/2)(\mathbf{B} + \mathbf{B}^T) = \operatorname{Re}(\mathbf{B})$. ■

We close this discussion by noting that, based on the results above, analytic approximations of the mean and the variance of the MI of the effective channel as a function of the physical channel and of the phase noise covariance matrix $\boldsymbol{\Sigma}_\varphi$ can be derived. Unlike the results in Section V, which are restricted to the (extreme) case of fully uncorrelated phase noise, these approximations can be obtained for general phase noise covariance matrices.

V. The Rank-1 Physical Channel

As already mentioned in Section III, the impact of phase errors is more pronounced for low-rank physical MIMO channels. In the following, we shall therefore analyze the extreme case of a rank-1 physical channel in detail. In practice, deterministic rank-1 channels occur in line-of-sight (LOS) scenarios with small angle-spread [3] (green-field like propagation conditions). Stochastic rank-1 MIMO channels are channels where the realization of the MIMO channel matrix has rank 1 w.p. 1. A prominent member of this class of channels is the pin-hole [3] or key-hole [4], [7], [8] channel reflecting propagation conditions with significant scattering close to the transmitter and the receiver and at the same time long distances between transmitter and receiver. Finally, the following MIMO channel sounder "calibration procedure" provides a practical motivation for studying (and quantifying) the impact of phase errors on rank-1 physical channels. The main idea underlying this calibration procedure is based on the fact that connecting transmitter and receiver in a TDMS-based sounder by a cable results in a deterministic rank-1 physical channel $\mathbf{H} = \alpha\mathbf{1}$, where $\alpha \in \mathbb{C}$ is the gain corresponding to the constant (across frequency) cable transfer function. The channel sounder then acquires samples of the effective channel matrix which contains channel coefficients that (after power normalization) have unit magnitude and a phase that varies due to phase errors created by the sounder. An inspection of the resulting eigenvalue histogram yields the number of significant eigenvalues and the corresponding eigenvalue distribution. Since the underlying physical channel has rank 1, it follows that any additional (w.r.t. the one resulting from the physical channel) significant modes in the effective channel must necessarily be due to phase errors (and/or potentially other imperfections in the measurement equipment). This "calibration measurement" can therefore loosely be interpreted as revealing the highest possible rank increase due to phase errors.

[3] A matrix $\mathbf{X}$ is said to be real-symmetric if $\mathbf{X} = \mathbf{X}^T$.

We shall see that, unlike the general case discussed in Sections III and IV, rank-1 physical channels allow to establish a number of insightful analytic results on the impact of phase errors on MI. We consider channels given by $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ where the vectors $\mathbf{g}$ and $\mathbf{h}$ can be either deterministic or stochastic (and with entries that are not necessarily unit modulus). The results in this section are valid for any $M_\mathrm{T}, M_\mathrm{R}$.

Let us start with a simple basic result which will be needed later in this section.

Lemma 1: The Hadamard product of a rank-1 matrix $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ and an arbitrary matrix $\boldsymbol{\Theta}$ can be written as a matrix product according to

$$(\mathbf{g}\mathbf{h}^T) \circ \boldsymbol{\Theta} = \mathrm{diag}(\mathbf{g})\, \boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h}).$$

Proof: The elements of a rank-1 matrix $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ are given by $[\mathbf{H}]_{m,n} = [\mathbf{g}]_m[\mathbf{h}]_n$. Consequently, we have $[\mathbf{H} \circ \boldsymbol{\Theta}]_{m,n} = [\mathbf{g}]_m[\mathbf{h}]_n[\boldsymbol{\Theta}]_{m,n}$. On the other hand, it follows immediately that $[\mathrm{diag}(\mathbf{g})\, \boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h})]_{m,n} = [\mathbf{g}]_m\, [\boldsymbol{\Theta}]_{m,n}\, [\mathbf{h}]_n$ which concludes the proof. ■

The following three Theorems state that a physical rank-1 channel subject to severe enough phase errors results in a full-rank effective channel.

Theorem 2: For a rank-1 physical channel $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ subject to phase errors with $\boldsymbol{\Theta}$ in (6) having full rank w.p. 1, we get

$$\det(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H) = \left(\prod_{i=1}^{M_\mathrm{R}} \left|[\mathbf{g}]_i\right|^2 \varepsilon_i\right) \det(\boldsymbol{\Theta}\boldsymbol{\Theta}^H), \quad M_\mathrm{R} \le M_\mathrm{T}$$

$$\det(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}) = \left(\prod_{i=1}^{M_\mathrm{T}} \left|[\mathbf{h}]_i\right|^2 \nu_i\right) \det(\boldsymbol{\Theta}^H\boldsymbol{\Theta}), \quad M_\mathrm{R} \ge M_\mathrm{T}$$

where

$$\min_i \left|[\mathbf{h}]_i\right|^2 \le \varepsilon_i \le \max_i \left|[\mathbf{h}]_i\right|^2,$$
$$\min_i \left|[\mathbf{g}]_i\right|^2 \le \nu_i \le \max_i \left|[\mathbf{g}]_i\right|^2.$$

Proof: We provide the proof for the case $M_\mathrm{R} \le M_\mathrm{T}$ only. The proof for $M_\mathrm{R} > M_\mathrm{T}$ follows exactly the same line of reasoning. We start by noting that Lemma 1 implies

$$\begin{aligned}
&\det(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H)\\
&\quad= \det\!\big(\mathrm{diag}(\mathbf{g})\, \boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h})\, \big(\mathrm{diag}(\mathbf{h})\big)^H \boldsymbol{\Theta}^H \big(\mathrm{diag}(\mathbf{g})\big)^H\big)\\
&\quad= \left(\prod_{i=1}^{M_\mathrm{R}} \left|[\mathbf{g}]_i\right|^2\right) \prod_{i=1}^{M_\mathrm{R}} \lambda_i\big(\boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h})\, \big(\mathrm{diag}(\mathbf{h})\big)^H \boldsymbol{\Theta}^H\big)\\
&\quad\overset{\text{(a)}}{=} \left(\prod_{i=1}^{M_\mathrm{R}} \left|[\mathbf{g}]_i\right|^2\right) \prod_{i=1}^{M_\mathrm{R}} \lambda_i\big(\big(\mathrm{diag}(\mathbf{h})\big)^H \boldsymbol{\Theta}^H \boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h})\big)\\
&\quad\overset{\text{(b)}}{=} \left(\prod_{i=1}^{M_\mathrm{R}} \left|[\mathbf{g}]_i\right|^2\right) \prod_{i=1}^{M_\mathrm{R}} \varepsilon_i\, \lambda_i\big(\boldsymbol{\Theta}\boldsymbol{\Theta}^H\big)
\end{aligned}$$

where the second product on the RHS of (a) is taken over the M_R nonzero eigenvalues of $\big(\mathrm{diag}(\mathbf{h})\big)^H \boldsymbol{\Theta}^H \boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h})$ only (note that the λ_i are ordered as defined in the Notations section) and (b) follows from a Corollary to Ostrowski's Theorem [18, Corollary 4.5.11]. ■

Theorem 2 thus states that a physical rank-1 channel subject to phase noise such that $\boldsymbol{\Theta}$ has full rank w.p. 1 results in a full-rank effective MIMO channel (provided that $[\mathbf{g}]_m \neq 0$, $\forall\, m$, and $[\mathbf{h}]_n \neq 0$, $\forall\, n$). For a deterministic physical rank-1 channel, the resulting effective channel will be stochastic and will have full rank w.p. 1. The condition of $\boldsymbol{\Theta}$ having full rank w.p. 1 may sound stringent. It turns out, however, that a full-rank phase noise covariance matrix $\mathrm{Cov}\{\boldsymbol{\Phi}\}$ is sufficient for $\boldsymbol{\Theta}$ to have full rank w.p. 1. This statement can be formalized as follows.

Lemma 2: A real Gaussian random matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M_\mathrm{R} \times M_\mathrm{T}}$ where $\mathrm{vec}(\boldsymbol{\Phi}) \overset{\mathrm{d}}{=} \mathcal{N}(\mathbf{0}, \mathrm{Cov}\{\boldsymbol{\Phi}\})$ with $\det(\mathrm{Cov}\{\boldsymbol{\Phi}\}) > 0$, has full rank w.p. 1. The matrix $\boldsymbol{\Theta} = \exp^\circ(j\boldsymbol{\Phi})$ has full rank w.p. 1 as well.

Proof: We follow the direct proof of [19, Th. 2.3, p. 712], where it is shown that for an $M_\mathrm{R} \times M_\mathrm{T}$ random matrix $\mathbf{X}$ to be full rank w.p. 1, it is sufficient to have the multivariate distribution of $\mathbf{X}$ be absolutely continuous w.r.t. $M_\mathrm{R}M_\mathrm{T}$-dimensional Lebesgue measure. This condition is trivially satisfied by $\boldsymbol{\Phi}$ with $\mathrm{vec}(\boldsymbol{\Phi}) \overset{\mathrm{d}}{=} \mathcal{N}(\mathbf{0}, \mathrm{Cov}\{\boldsymbol{\Phi}\})$ and $\det(\mathrm{Cov}\{\boldsymbol{\Phi}\}) > 0$ (see, e.g., [20, Sec. 4.7.2]).

The second part of the statement can be proved by using [21, Lemma 3], which states that $\boldsymbol{\Theta}$ has full rank if either $\mathrm{Re}(\boldsymbol{\Theta})$, $\mathrm{Im}(\boldsymbol{\Theta})$, or $\left[\,\big(\mathrm{Re}(\boldsymbol{\Theta})\big)^T \quad \big(\mathrm{Im}(\boldsymbol{\Theta})\big)^T\right]^T$ has full rank. Direct computation reveals that the multivariate pdf of $\big(\mathrm{Re}(\boldsymbol{\Theta})\big)^T = \cos^\circ(\boldsymbol{\Theta})$ is continuous and integrable (in the interval $[-1, 1]^{M_\mathrm{T}M_\mathrm{R}}$) so that its multivariate cdf is absolutely continuous w.r.t. $M_\mathrm{R}M_\mathrm{T}$-dimensional Lebesgue measure. Hence, by the direct proof of [19, Th. 2.3], $\mathrm{Re}(\boldsymbol{\Theta})$ is full rank w.p. 1, and from what was said before it follows that $\boldsymbol{\Theta}$ has full rank w.p. 1. ■

Besides what was stated in Theorem 2 above, relating properties of $\mathrm{Cov}\{\boldsymbol{\Phi}\}$ to properties of $\det(\boldsymbol{\Theta}\boldsymbol{\Theta}^H)$ seems difficult. For $M_\mathrm{T} = M_\mathrm{R}$, we can refine the result in Theorem 2 as follows.

Theorem 3: For a rank-1 physical channel $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ with $M_\mathrm{T} = M_\mathrm{R} = M$ subject to phase errors with $\boldsymbol{\Theta}$ having full rank w.p. 1, we have

$$\det(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H) = \left(\prod_{i=1}^{M} \left|[\mathbf{h}]_i\right|^2\right) \left(\prod_{i=1}^{M} \left|[\mathbf{g}]_i\right|^2\right) \det(\boldsymbol{\Theta}\boldsymbol{\Theta}^H). \tag{29}$$

Proof: The proof follows trivially using Lemma 1 and noting that

$$\begin{aligned}
&\det(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H)\\
&\quad= \det\Big(\mathrm{diag}(\mathbf{g})\, \boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h})\, \big(\mathrm{diag}(\mathbf{h})\big)^H \boldsymbol{\Theta}^H \big(\mathrm{diag}(\mathbf{g})\big)^H\Big)\\
&\quad= \det\Big(\big(\mathrm{diag}(\mathbf{g})\big)^H \mathrm{diag}(\mathbf{g})\Big)\\
&\qquad\qquad \cdot \det\Big(\boldsymbol{\Theta}\, \mathrm{diag}(\mathbf{h})\, \big(\mathrm{diag}(\mathbf{h})\big)^H \boldsymbol{\Theta}^H\Big)\\
&\quad= \det\Big(\big(\mathrm{diag}(\mathbf{g})\big)^H \mathrm{diag}(\mathbf{g})\Big)\\
&\qquad\qquad \cdot \det\Big(\mathrm{diag}(\mathbf{h})\, \big(\mathrm{diag}(\mathbf{h})\big)^H\Big) \det\big(\boldsymbol{\Theta}\boldsymbol{\Theta}^H\big)
\end{aligned}$$

which yields (29). ■

The following Theorem allows a more specific conclusion since it shows that for rank-1 channels $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ where $\mathbf{g}$ and $\mathbf{h}$ consist of unit-modulus entries (representative of LOS propagation [3]) the rank of the effective channel matrix is equal to the rank of $\boldsymbol{\Theta}$. Moreover, the eigenvalues of the effective channel matrix (more specifically of $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H$) are equal to the eigenvalues of $\boldsymbol{\Theta}\boldsymbol{\Theta}^H$.

Theorem 4: For a rank-1 physical channel $\mathbf{H} = \mathbf{g}\mathbf{h}^T$, where $\mathbf{g}$ and $\mathbf{h}$ are such that $\left|[\mathbf{g}]_i\right| = 1$ ($i = 1, 2, \ldots, M_\mathrm{R}$) and $\left|[\mathbf{h}]_i\right| = 1$ ($i = 1, 2, \ldots, M_\mathrm{T}$), we have

$$\lambda_i(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H) = \lambda_i(\boldsymbol{\Theta}\boldsymbol{\Theta}^H), \quad i = 1, 2, \ldots, M_\mathrm{R}, \quad M_\mathrm{R} \leq M_\mathrm{T}$$
$$\lambda_i(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}) = \lambda_i(\boldsymbol{\Theta}^H\boldsymbol{\Theta}), \quad i = 1, 2, \ldots, M_\mathrm{T}, \quad M_\mathrm{R} > M_\mathrm{T}.$$

Proof: The proof for both cases is trivially obtained using Lemma 1 and noting that the assumptions of the Theorem imply $\mathrm{diag}(\mathbf{g})\left(\mathrm{diag}(\mathbf{g})\right)^H = \mathbf{I}_{M_\mathrm{R}}$ and $\mathrm{diag}(\mathbf{h})\left(\mathrm{diag}(\mathbf{h})\right)^H = \mathbf{I}_{M_\mathrm{T}}$. For $M_\mathrm{R} \leq M_\mathrm{T}$, simply note that

$$\begin{aligned}
&\lambda_i(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H)\\
&\quad = \lambda_i\Big(\mathrm{diag}(\mathbf{g})\,\boldsymbol{\Theta}\,\mathrm{diag}(\mathbf{h})\left(\mathrm{diag}(\mathbf{h})\right)^H\boldsymbol{\Theta}^H\left(\mathrm{diag}(\mathbf{g})\right)^H\Big)\\
&\quad = \lambda_i\Big(\left(\mathrm{diag}(\mathbf{g})\right)^H\mathrm{diag}(\mathbf{g})\,\boldsymbol{\Theta}\boldsymbol{\Theta}^H\Big)\\
&\quad = \lambda_i(\boldsymbol{\Theta}\boldsymbol{\Theta}^H), \quad i = 1, 2, \ldots, M_\mathrm{R}.
\end{aligned}$$

The case $M_\mathrm{R} > M_\mathrm{T}$ follows exactly the same line of reasoning. ∎

Since the high-SNR MI of $\widehat{\mathbf{H}}$ (for $M_\mathrm{R} \leq M_\mathrm{T}$) is given by $\hat{I} \approx \log\det\left((\rho/M_\mathrm{T})\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H\right)$, Theorems 2 and 3 immediately yield expressions[4] for the high-SNR MI of $\widehat{\mathbf{H}}$. However, the pdf of the quantity $\log\det(\boldsymbol{\Theta}\boldsymbol{\Theta}^H)$ is, in general, difficult to obtain. Insightful analytic results are, however, possible by invoking the small phase noise assumption $\exp^\circ(j\boldsymbol{\Phi}) \approx \mathbf{1} + j\boldsymbol{\Phi}$. As demonstrated previously, the small phase noise approximation is very well satisfied in practice as the worst-case value of 7° root mean-square (rms) phase noise amounts to $\sigma_\varphi^2 \approx 0.0149$. Therefore, as a consequence of i) the small phase noise approximation and ii) the frequency offset having no effect due to the regular sounding pattern, it suffices to analyze the quantity $\det(\widetilde{\boldsymbol{\Theta}}\widetilde{\boldsymbol{\Theta}}^H)$ with $\widetilde{\boldsymbol{\Theta}} = \mathbf{1} + j\boldsymbol{\Phi}$ instead of $\det(\boldsymbol{\Theta}\boldsymbol{\Theta}^H)$. Interestingly, $\det(\widetilde{\boldsymbol{\Theta}}\widetilde{\boldsymbol{\Theta}}^H)$ can be characterized in terms of chi-square RVs and a beta-distributed RV, which provides the basis for tight bounds on $\widehat{C} = \mathbb{E}\{\hat{I}\}$ and for accurate approximations of $\mathrm{Var}\{\hat{I}\}$. Before stating the corresponding results using the exact expression for $\det(\widetilde{\boldsymbol{\Theta}}\widetilde{\boldsymbol{\Theta}}^H)$, we shall, however, provide an approximation for $\det(\widetilde{\boldsymbol{\Theta}}\widetilde{\boldsymbol{\Theta}}^H)$ (in the sense of distributional equivalence), which turns out to be particularly useful to derive a simple analytic lower bound on $\widehat{C}$ (see Theorem 8). This approximation is based on the following result.

Theorem 5: Under the small phase noise approximation $\sigma_\varphi^2 \ll 1$ so that $\widetilde{\boldsymbol{\Theta}} = \mathbf{1} + j\boldsymbol{\Phi}$, assuming fully uncorrelated phase noise, i.e., $\mathrm{vec}(\boldsymbol{\Phi}) \overset{\mathrm{d}}{=} \mathcal{N}(\mathbf{0}, \sigma_\varphi^2\mathbf{I}_{M_\mathrm{T}M_\mathrm{R}})$, we have

$$\begin{aligned}
&\det(\widetilde{\boldsymbol{\Theta}}\widetilde{\boldsymbol{\Theta}}^H)\\
&\quad \overset{\mathrm{d}}{=} \left(\chi^2_{M_\mathrm{T},\sigma_\varphi^2} + M_\mathrm{T}M_\mathrm{R}\right)\prod_{i=2}^{M_\mathrm{R}}\left(\chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2} + Z(\eta^{(i)})\right),\\
&\qquad\qquad\qquad\qquad\qquad\qquad M_\mathrm{R} \leq M_\mathrm{T}
\end{aligned} \tag{30a}$$

$$\begin{aligned}
&\det(\widetilde{\boldsymbol{\Theta}}^H\widetilde{\boldsymbol{\Theta}})\\
&\quad \overset{\mathrm{d}}{=} \left(\chi^2_{M_\mathrm{R},\sigma_\varphi^2} + M_\mathrm{T}M_\mathrm{R}\right)\prod_{i=2}^{M_\mathrm{T}}\left(\chi^2_{M_\mathrm{R}-i,\sigma_\varphi^2} + Z(\eta^{(i)})\right),\\
&\qquad\qquad\qquad\qquad\qquad\qquad M_\mathrm{R} > M_\mathrm{T}
\end{aligned} \tag{30b}$$

where the χ^2_{n,σ^2} are statistically independent[5] and $Z(\eta^{(i)}) = \sigma_\varphi^2(\eta^{(i)}X_1^{(i)} + (1-\eta^{(i)})X_2^{(i)})$ with $X_1^{(i)}, X_2^{(i)}$ i.i.d. $\chi^2_{1,1}$ and the $\eta^{(i)}$, $\forall\, i$, being RVs with pdf supported in the interval [0,1].

Proof: We provide the proof for $M_\mathrm{R} \leq M_\mathrm{T}$ only. The case $M_\mathrm{R} > M_\mathrm{T}$ follows exactly the same line of reasoning. Let us start by noting that the singular value decomposition of $\mathbf{1}_{M_\mathrm{R},M_\mathrm{T}}$ is given by $\mathbf{1}_{M_\mathrm{R},M_\mathrm{T}} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{W}^T$, where $\mathbf{V}$ is of dimension $M_\mathrm{R} \times M_\mathrm{R}$, $\mathbf{W}$ is $M_\mathrm{T} \times M_\mathrm{T}$, and the $M_\mathrm{R} \times M_\mathrm{T}$ matrix $\boldsymbol{\Sigma}$ is given by

$$[\boldsymbol{\Sigma}]_{m,n} = \begin{cases} \sqrt{M_\mathrm{T}M_\mathrm{R}}, & m = n = 1\\ 0, & \text{else.}\end{cases} \tag{31}$$

Defining the $M_\mathrm{R} \times M_\mathrm{T}$ matrix $\mathbf{S} = -j\boldsymbol{\Sigma} + \widetilde{\boldsymbol{\Phi}}$ with $\widetilde{\boldsymbol{\Phi}} = \mathbf{V}^T\boldsymbol{\Phi}\mathbf{W}$ (and hence $\widetilde{\boldsymbol{\Phi}} \overset{\mathrm{d}}{=} \boldsymbol{\Phi}$), it follows that $\det(\widetilde{\boldsymbol{\Theta}}\widetilde{\boldsymbol{\Theta}}^H) = \det(\mathbf{S}\mathbf{S}^H)$. With $\mathbf{S} = [\,\mathbf{s}_1 \;\; \mathbf{s}_2 \;\; \cdots \;\; \mathbf{s}_{M_\mathrm{R}}\,]^T$ being a square ($M_\mathrm{R} = M_\mathrm{T}$) or a wide matrix ($M_\mathrm{R} < M_\mathrm{T}$), a basic result in geometry (e.g., [22, Th. 7.5.1], [23, Sec. 3.2.2], which can be shown to hold in the complex case upon replacing transposition by conjugate transposition) yields

$$\sqrt{\det(\mathbf{S}\mathbf{S}^H)} = \mathrm{vol}(P_\mathbf{S}) = \|\mathbf{s}_1^\perp\|\,\|\mathbf{s}_2^\perp\|\cdots\|\mathbf{s}_{M_\mathrm{R}}^\perp\| \tag{32}$$

where $\mathrm{vol}(P_\mathbf{S})$ stands for the volume or M_R-content of the parallelotope spanned by the M_R row vectors of $\mathbf{S}$, $\mathbf{s}_1^\perp = \mathbf{s}_1$, and $\mathbf{s}_i^\perp$ ($i > 1$) denotes the component of $\mathbf{s}_i$ orthogonal to the span of the vectors $\mathbf{s}_1^\perp, \mathbf{s}_2^\perp, \ldots, \mathbf{s}_{i-1}^\perp$. The orthogonal vectors $\mathbf{s}_i^\perp$ ($i = 2, 3, \ldots, M_\mathrm{R}$) are obtained using Gram-Schmidt orthogonalization and are given by

$$\mathbf{s}_i^\perp = \left(\mathbf{I}_{M_\mathrm{T}} - \sum_{n=1}^{i-1}\frac{\mathbf{s}_n^\perp\mathbf{s}_n^{\perp H}}{\|\mathbf{s}_n^\perp\|^2}\right)\mathbf{s}_i = \mathbf{A}_i\mathbf{s}_i. \tag{33}$$

It is well known that applying the decomposition (32) to an i.i.d. complex Gaussian random matrix $\mathbf{S}$ with $\mathcal{CN}(0,1)$ elements results in independent chi-square distributed factors $\|\mathbf{s}_i^\perp\|^2$ ($i = 1, 2, \ldots, M_\mathrm{R}$) [24, Th. 3.4 ff.]. The problem at hand differs, however, from the i.i.d. complex Gaussian case in two aspects, namely the fact that the elements in $\boldsymbol{\Phi}$ and hence $\widetilde{\boldsymbol{\Phi}}$ are real-valued Gaussian and the presence of the deterministic component $-j\boldsymbol{\Sigma}$.

It follows trivially from the definition of $\mathbf{S}$ that $\|\mathbf{s}_1^\perp\|^2 \overset{\mathrm{d}}{=} \chi^2_{M_\mathrm{T},\sigma_\varphi^2} + M_\mathrm{T}M_\mathrm{R}$. From (33) we can see that, conditioned on $\mathbf{s}_1^\perp, \mathbf{s}_2^\perp, \ldots, \mathbf{s}_{i-1}^\perp$, the vectors $\mathbf{s}_i^\perp$ ($i = 2, 3, \ldots, M_\mathrm{R}$) are JG and hence the $\|\mathbf{s}_i^\perp\|^2$ ($i = 2, 3, \ldots, M_\mathrm{R}$) are chi-square distributed. Using the fact that $\mathbf{s}_i \in \mathbb{R}^{M_\mathrm{T}}$ ($i = 2, 3, \ldots, M_\mathrm{R}$) and $\mathbf{A}_i^H\mathbf{A}_i = \mathbf{A}_i$, it follows immediately that

$$\|\mathbf{s}_i^\perp\|^2 = \mathbf{s}_i^T\mathbf{A}_i\mathbf{s}_i, \quad i = 2, 3, \ldots, M_\mathrm{R}.$$

[4]More specifically an approximation in the case of Theorem 2 due to the presence of the quantities ε_i and ν_i.

[5]Note that the product over i on the RHS of (30) is equal to 1 if $M_\mathrm{R} = 1$ (in the case $M_\mathrm{R} \leq M_\mathrm{T}$) and $M_\mathrm{T} = 1$ (in the case $M_\mathrm{R} > M_\mathrm{T}$).

Next, noting that

$$\|\mathbf{s}_i^{\perp}\|^2 = \mathbf{s}_i^T\left(\mathrm{Re}(\mathbf{A}_i) + j\mathrm{Im}(\mathbf{A}_i)\right)\mathbf{s}_i = \mathbf{s}_i^T\,\mathrm{Re}(\mathbf{A}_i)\,\mathbf{s}_i + j\mathbf{s}_i^T\,\mathrm{Im}(\mathbf{A}_i)\,\mathbf{s}_i, \quad i = 2,3,\dots,M_\mathrm{R}$$

has to be real-valued for all $\mathbf{s}_i$, it follows that

$$\|\mathbf{s}_i^{\perp}\|^2 = \mathbf{s}_i^T\,\mathrm{Re}(\mathbf{A}_i)\,\mathbf{s}_i = \mathbf{s}_i^T\left(\mathbf{I}_{M_\mathrm{T}} - \sum_{n=1}^{i-1}\frac{\mathrm{Re}(\mathbf{s}_n^{\perp}\mathbf{s}_n^{\perp H})}{\|\mathbf{s}_n^{\perp}\|^2}\right)\mathbf{s}_i. \tag{34}$$

Based on (34), we can now invoke Lemma 4 in the Appendix to conclude that the eigenvalues of $\mathrm{Re}(\mathbf{A}_i)$ are given by

$$\{\sigma_k^{(i)}\} = \{\underbrace{1,\dots,1}_{M_\mathrm{T}-i},\underbrace{0,\dots,0}_{i-2},\eta^{(i)},1-\eta^{(i)}\}, \quad k = 1,2,\dots,M_\mathrm{T}$$

where $\eta^{(i)} = \eta^{(i)}(\mathbf{s}_1^{\perp},\mathbf{s}_2^{\perp},\dots,\mathbf{s}_{i-1}^{\perp})$ is a RV with pdf supported in the interval $[0,1]$. Consequently, using [25, Eq. (4.1.1)], we obtain

$$\mathbf{s}_i^T\,\mathrm{Re}(\mathbf{A}_i)\,\mathbf{s}_i \overset{\mathrm{d}}{=} \sigma_\varphi^2\sum_{k=1}^{M_\mathrm{T}}\sigma_k^{(i)}X_i = \chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2} + \underbrace{\sigma_\varphi^2(\eta^{(i)}X_{M_\mathrm{T}-1}^{(i)} + (1-\eta^{(i)})X_{M_\mathrm{T}}^{(i)})}_{\overset{\mathrm{d}}{=}\ Z(\eta^{(i)})}$$

where the $X_i \overset{\mathrm{d}}{=} \chi^2_{1,1}$ are independent. ■

We shall next show that for $\sigma_\varphi^2 \ll 1$, $Z(\eta^{(i)}) \overset{\mathrm{d}}{\approx} \chi^2_{1,\sigma_\varphi^2}$, which then implies that

$$\|\mathbf{s}_i^{\perp}\|^2 \overset{\mathrm{d}}{\approx} \chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2} + \chi^2_{1,\sigma_\varphi^2} = \chi^2_{M_\mathrm{T}-i+1,\sigma_\varphi^2} \tag{35}$$

thereby allowing an approximation of (30) as[6]

$$\begin{aligned}\det(\widetilde{\mathbf{\Theta}}\widetilde{\mathbf{\Theta}}^H) &\overset{\mathrm{d}}{\approx} \left(\chi^2_{M_\mathrm{T},\sigma_\varphi^2} + M_\mathrm{T}M_\mathrm{R}\right)\prod_{i=1}^{M_\mathrm{R}-1}\chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2}, \quad M_\mathrm{R} \le M_\mathrm{T}\\ \det(\widetilde{\mathbf{\Theta}}^H\widetilde{\mathbf{\Theta}}) &\overset{\mathrm{d}}{\approx} \left(\chi^2_{M_\mathrm{R},\sigma_\varphi^2} + M_\mathrm{T}M_\mathrm{R}\right)\prod_{i=1}^{M_\mathrm{T}-1}\chi^2_{M_\mathrm{R}-i,\sigma_\varphi^2}, \quad M_\mathrm{R} > M_\mathrm{T}.\end{aligned} \tag{36}$$

In order to see that $Z(\eta^{(i)}) \overset{\mathrm{d}}{\approx} \chi^2_{1,\sigma_\varphi^2}$, we start by noting that the pdf of $Z(\eta)$ conditional on η is given by [27, Eq. (5.7)]

$$\mathrm{p}_{Z|\eta}(x) = \frac{1}{2\sigma_\varphi^2\sqrt{\eta(1-\eta)}}\,e^{-\frac{x}{4\sigma_\varphi^2\eta(1-\eta)}}\,I_0\!\left(\frac{1-2\eta}{4\sigma_\varphi^2\eta(1-\eta)}\,x\right) \tag{37}$$

where $I_0(z)$ is the modified Bessel function of the first kind [11, Sec. 9.6]. For σ_φ^2 small, we can invoke the large-$|z|$ expansion of $I_0(z)$ [11, Eq. 9.7.1] according to

$$I_0(z) = \frac{1}{\sqrt{2\pi z}}\,e^z\left(1 + \frac{1}{8z} + \frac{3^2}{2!(8z)^2} + \frac{3^2 5^2}{3!(8z)^3} + \cdots\right) \approx \frac{1}{\sqrt{2\pi z}}\,e^z$$

which, when used in (37), upon renormalizing so that $\int_{x=0}^{\infty}\mathrm{p}_{Z|\eta}(x)\,\mathrm{d}x = 1$, yields

$$\mathrm{p}_{Z|\eta}(x) \approx \frac{1}{\sqrt{2\pi\sigma_\varphi^2(1-\eta)x}}\,e^{-\frac{x}{2\sigma_\varphi^2(1-\eta)}} = \mathrm{p}_{\chi^2_{1,\sigma_\varphi^2(1-\eta)}}(x).$$

This means that $Z|\eta \overset{\mathrm{d}}{=} \chi^2_{1,\sigma_\varphi^2(1-\eta)}$ for $0 < \eta < 1$ if σ_φ^2 is small.

We shall next see that $\eta^{(i)}, \forall i$, is small, in general, which then directly results in the (unconditional) pdf of $Z(\eta^{(i)})$ satisfying $Z(\eta^{(i)}) \overset{\mathrm{d}}{\approx} \chi^2_{1,\sigma_\varphi^2}, \forall i$. Recall that $\{\eta^{(i)}, 1-\eta^{(i)}\}$ are the eigenvalues of $\mathrm{Re}(\mathbf{A}_i)$ in (34) that equal neither zero nor one. The first pair of such eigenvalues is obtained for $i = 2$. Due to the symmetry of the eigenvalues, we may investigate $\mathbf{I} - \mathrm{Re}(\mathbf{A}_2)$ instead of $\mathrm{Re}(\mathbf{A}_2)$, which, using $\mathbf{s}_1^{\perp} = \mathbf{s}_1$, can be written as

$$\mathbf{I} - \mathrm{Re}(\mathbf{A}_2) = \frac{\mathrm{Re}(\mathbf{s}_1\mathbf{s}_1^H)}{\|\mathbf{s}_1\|^2} = \frac{\mathrm{Re}(\mathbf{s}_1)\,(\mathrm{Re}(\mathbf{s}_1))^T}{\|\mathbf{s}_1\|^2} + \frac{\mathrm{Im}(\mathbf{s}_1)\,(\mathrm{Im}(\mathbf{s}_1))^T}{\|\mathbf{s}_1\|^2} = \frac{\boldsymbol{\varphi}\boldsymbol{\varphi}^T}{\|\boldsymbol{\varphi}\|^2 + \|\boldsymbol{\sigma}\|^2} + \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}^T}{\|\boldsymbol{\varphi}\|^2 + \|\boldsymbol{\sigma}\|^2}$$

where we set $\boldsymbol{\varphi} = \mathrm{Re}(\mathbf{s}_1)$ and $\boldsymbol{\sigma} = -\mathrm{Im}(\mathbf{s}_1) = [\sqrt{M_\mathrm{T}M_\mathrm{R}}\ \ 0\ \ 0\ \ \cdots\ \ 0]^T$. In the following, we denote $\varphi_i = [\boldsymbol{\varphi}]_i$. The nonzero eigenvalues of $\mathbf{I} - \mathrm{Re}(\mathbf{A}_2)$ are equal to the eigenvalues of

$$\frac{[\boldsymbol{\varphi}\ \ \boldsymbol{\sigma}]^T[\boldsymbol{\varphi}\ \ \boldsymbol{\sigma}]}{\|\boldsymbol{\varphi}\|^2 + \|\boldsymbol{\sigma}\|^2} = \frac{1}{\|\boldsymbol{\varphi}\|^2 + \|\boldsymbol{\sigma}\|^2}\begin{bmatrix}\|\boldsymbol{\varphi}\|^2 & \sqrt{M_\mathrm{T}M_\mathrm{R}}\,\varphi_1\\ \sqrt{M_\mathrm{T}M_\mathrm{R}}\,\varphi_1 & M_\mathrm{T}M_\mathrm{R}\end{bmatrix}$$

given by

$$\{\eta^{(2)}, 1-\eta^{(2)}\} = \frac{1}{2} \pm \frac{1}{2}\frac{\sqrt{4M_\mathrm{T}M_\mathrm{R}\varphi_1^2 + (M_\mathrm{T}M_\mathrm{R} - \|\boldsymbol{\varphi}\|^2)^2}}{M_\mathrm{T}M_\mathrm{R} + \|\boldsymbol{\varphi}\|^2}. \tag{38}$$

For M_T sufficiently large, with $\|\boldsymbol{\varphi}\|^2 = \varphi_1^2 + \varphi_\mathrm{s}^2$, where $\varphi_\mathrm{s}^2 = \sum_{i=2}^{M_\mathrm{T}}\varphi_i^2$, we can replace (38) by

$$\{\eta^{(2)}, 1-\eta^{(2)}\} \approx \frac{1}{2} \pm \frac{1}{2}\frac{M_\mathrm{T}M_\mathrm{R} - \varphi_\mathrm{s}^2}{M_\mathrm{T}M_\mathrm{R} + \varphi_\mathrm{s}^2}.$$

Next, since φ_s^2 is small compared to $M_\mathrm{T}M_\mathrm{R}$, we obtain the first-order Taylor series expansion

$$\{\eta^{(2)}, 1-\eta^{(2)}\} \approx \frac{1}{2} \pm \frac{1}{2}\left(1 - 2\frac{\varphi_\mathrm{s}^2}{M_\mathrm{T}M_\mathrm{R}}\right) \overset{\mathrm{d}}{=} \left\{\chi^2_{M_\mathrm{T}-1,\sigma_\varphi^2/(M_\mathrm{T}M_\mathrm{R})},\ 1 - \chi^2_{M_\mathrm{T}-1,\sigma_\varphi^2/(M_\mathrm{T}M_\mathrm{R})}\right\}.$$

[6]We would like to use this chance to point out that the distributional equivalence in [26, Prop. 4] should be an approximate equivalence (as in (36)). Furthermore, $\mathcal{CN}(\mathbf{0},\sigma_\Phi^2\mathbf{I}_{M_\mathrm{T}M_\mathrm{R}})$ in [26, Prop. 4 and Prop. 5] should be replaced by $\mathcal{N}(\mathbf{0},\sigma_\Phi^2\mathbf{I}_{M_\mathrm{T}M_\mathrm{R}})$.

Hence, we have

$$\mathbb{E}\{\eta^{(2)}\} = \sigma_\varphi^2 \frac{M_\mathrm{T}-1}{M_\mathrm{T}M_\mathrm{R}} \qquad \mathrm{Var}\{\eta^{(2)}\} = 2\sigma_\varphi^4 \frac{M_\mathrm{T}-1}{M_\mathrm{T}^2 M_\mathrm{R}^2}$$

which shows that, for sufficiently large $M_\mathrm{T}, M_\mathrm{R}$, $\eta^{(2)}$ is indeed small. For $M_\mathrm{R} \times 1$ and $1 \times M_\mathrm{T}$ systems, i.e., for single-input multiple-output (SIMO) and multiple-input single-output (MISO) systems, respectively, we can therefore immediately conclude that the approximation (36) is very accurate. In the case of general M_T and M_R, it seems difficult to prove that $\eta^{(i)} \approx 0$ for $i \geq 3$. We do, however, have strong numerical evidence that this is, indeed, the case.

Recalling that $\sigma_\varphi^2 \approx 0.0149$ for the worst-case phase noise value of 7° rms, we can conclude that the assumption $\sigma_\varphi^2 \ll 1$ made in Theorem 5 and in (36) is very well satisfied in practice. Fig. 2 shows the cdf of $\log\det(\widetilde{\mathbf{\Theta}}\widetilde{\mathbf{\Theta}}^H)$ corresponding to the approximation (36) along with the exact cdf[7] (in both cases obtained through Monte Carlo methods). We observe that the approximation is excellent in general and, indeed, becomes better for smaller σ_φ^2 and/or for less symmetric (in terms of the number of transmit and receive antennas) configurations. We finally note that comparing (36) to [28, Eq. (3)] suggests that for fully uncorrelated phase noise with $\sigma_\varphi^2 \ll 1$, the effective MIMO channel behaves like a physical MIMO channel consisting of a rank-1 Ricean component plus an i.i.d. Rayleigh fading component with the difference that in our case the chi-square RVs have half the order of those in [28] (reflecting the fact that here we are dealing with real-valued Gaussian RVs).

We proceed by stating the result on the exact distribution of $\det(\widetilde{\mathbf{\Theta}}\widetilde{\mathbf{\Theta}}^H)$.

Theorem 6: Under the small phase noise approximation $\sigma_\varphi^2 \ll 1$ so that $\widetilde{\mathbf{\Theta}} = \mathbf{1} + j\mathbf{\Phi}$, assuming fully uncorrelated phase noise, i.e., $\mathrm{vec}(\mathbf{\Phi}) \stackrel{\mathrm{d}}{=} \mathcal{N}(\mathbf{0}, \sigma_\varphi^2 \mathbf{I}_{M_\mathrm{T}M_\mathrm{R}})$, we have

$$\begin{aligned}
&\det(\widetilde{\mathbf{\Theta}}\widetilde{\mathbf{\Theta}}^H) \\
&\quad \stackrel{\mathrm{d}}{=} \left(\chi^2_{M_\mathrm{d}+1,\sigma_\varphi^2} + M_\mathrm{p}\, \beta\!\left(\frac{M_\mathrm{d}+1}{2}, \frac{M_\mathrm{R}-1}{2}\right)\right) \\
&\qquad\qquad \cdot \prod_{i=1}^{M_\mathrm{R}-1} \chi^2_{M_\mathrm{T}-i+1,\sigma_\varphi^2}, \\
&\qquad\qquad\qquad\qquad\qquad M_\mathrm{R} \leq M_\mathrm{T} \\
&\det(\widetilde{\mathbf{\Theta}}^H\widetilde{\mathbf{\Theta}}) \\
&\quad \stackrel{\mathrm{d}}{=} \left(\chi^2_{M_\mathrm{d}+1,\sigma_\varphi^2} + M_\mathrm{p}\, \beta\!\left(\frac{M_\mathrm{d}+1}{2}, \frac{M_\mathrm{T}-1}{2}\right)\right) \\
&\qquad\qquad \cdot \prod_{i=1}^{M_\mathrm{T}-1} \chi^2_{M_\mathrm{R}-i+1,\sigma_\varphi^2}, \\
&\qquad\qquad\qquad\qquad\qquad M_\mathrm{R} > M_\mathrm{T}
\end{aligned} \tag{39}$$

where $M_\mathrm{d} = |M_\mathrm{T} - M_\mathrm{R}|$, $M_\mathrm{p} = M_\mathrm{T}M_\mathrm{R}$, and the χ^2_{n,σ^2} are statistically independent.

Proof: Again, we provide the proof for $M_\mathrm{R} \leq M_\mathrm{T}$ only. The case $M_\mathrm{R} > M_\mathrm{T}$ follows exactly the same line of reasoning.

[7]Note that "exact cdf" means exact under the linear phase noise approximation.

We start by noting that the matrix $\mathbf{S}$ defined in the proof of Theorem 5 is unitarily equivalent to the matrix $\mathbf{S}' = -j\mathbf{\Sigma}' + \widetilde{\mathbf{\Phi}}'$, where the $M_\mathrm{R} \times M_\mathrm{T}$ matrix $\mathbf{\Sigma}'$ is given by

$$[\mathbf{\Sigma}']_{m,n} = \begin{cases} \sqrt{M_\mathrm{T}M_\mathrm{R}}, & m = n = M_\mathrm{R} \\ 0, & \text{else} \end{cases}$$

and $\widetilde{\mathbf{\Phi}}' \stackrel{\mathrm{d}}{=} \widetilde{\mathbf{\Phi}} \stackrel{\mathrm{d}}{=} \mathbf{\Phi}$. In what follows, we shall work with $\mathbf{S}'$ and, by slight abuse of notation, denote it as $\mathbf{S}$. The pdf of $\|\mathbf{s}_1^\perp\|^2$ follows trivially from the definition of $\mathbf{S}$ and is given by $\chi^2_{M_\mathrm{T},\sigma_\varphi^2}$. Applying Gram-Schmidt orthogonalization, due to the nonzero entry in $\mathbf{\Sigma}'$ being at position $m = n = M_\mathrm{R}$, we can conclude that for $i = 2, 3, \ldots, M_\mathrm{R}-1$, the matrix $\mathrm{Re}(\mathbf{A}_i)$ has only two distinct eigenvalues, namely 0 with multiplicity $i-1$ and 1 with multiplicity $M_\mathrm{T} - i + 1$. Consequently, (34) implies that $\|\mathbf{s}_i^\perp\|^2$ $(i = 1, 2, \ldots, M_\mathrm{R}-1)$ conditioned on $\mathbf{s}_1^\perp, \mathbf{s}_2^\perp, \ldots, \mathbf{s}_{i-1}^\perp$ is distributed as $\chi^2_{M_\mathrm{T}-i+1,\sigma_\varphi^2}$. Since the eigenvalues of $\mathrm{Re}(\mathbf{A}_i)$ do not depend on the vectors $\mathbf{s}_1^\perp, \mathbf{s}_2^\perp, \ldots, \mathbf{s}_{i-1}^\perp$ and the statistics of $\|\mathbf{s}_i^\perp\|^2$ depend on $\mathbf{A}_i$ only through the eigenvalues of $\mathbf{A}_i$, we can conclude that the unconditional distribution of $\|\mathbf{s}_i^\perp\|^2$ satisfies $\|\mathbf{s}_i^\perp\|^2 \stackrel{\mathrm{d}}{=} \chi^2_{M_\mathrm{T}-i+1,\sigma_\varphi^2}$ $(i = 1, 2, \ldots, M_\mathrm{R} - 1)$. For $i = M_\mathrm{R}$, noting that $\mathbf{A}_{M_\mathrm{R}}$ is a real-valued matrix, we have

$$\|\mathbf{s}_{M_\mathrm{R}}^\perp\|^2 = \|\mathbf{A}_{M_\mathrm{R}}\mathrm{Re}(\mathbf{s}_{M_\mathrm{R}})\|^2 + \|\mathbf{A}_{M_\mathrm{R}}\mathrm{Im}(\mathbf{s}_{M_\mathrm{R}})\|^2. \tag{40}$$

The distribution of the first term on the RHS of (40) can be shown, using the same line of reasoning as for $i = 1, 2, \ldots, M_\mathrm{R} - 1$, to satisfy $\|\mathbf{A}_{M_\mathrm{R}}\mathrm{Re}(\mathbf{s}_{M_\mathrm{R}})\|^2 \stackrel{\mathrm{d}}{=} \chi^2_{M_\mathrm{T}-M_\mathrm{R}+1,\sigma_\varphi^2}$. The second term on the RHS of (40) can be expanded as

$$\begin{aligned}
&\|\mathbf{A}_{M_\mathrm{R}}\mathrm{Im}(\mathbf{s}_{M_\mathrm{R}})\|^2 \\
&\quad = \big(\mathrm{Im}(\mathbf{s}_{M_\mathrm{R}})\big)^T \mathbf{A}_{M_\mathrm{R}}^T \mathbf{A}_{M_\mathrm{R}}\, \mathrm{Im}(\mathbf{s}_{M_\mathrm{R}}) \\
&\quad = \big(\mathrm{Im}(\mathbf{s}_{M_\mathrm{R}})\big)^T \mathbf{A}_{M_\mathrm{R}}\, \mathrm{Im}(\mathbf{s}_{M_\mathrm{R}})
\end{aligned}$$

where we made use of the fact that $\mathbf{A}_{M_\mathrm{R}}$ is real-valued and hence $\mathbf{A}_{M_\mathrm{R}}^H \mathbf{A}_{M_\mathrm{R}} = \mathbf{A}_{M_\mathrm{R}}$ reduces to $\mathbf{A}_{M_\mathrm{R}}^T \mathbf{A}_{M_\mathrm{R}} = \mathbf{A}_{M_\mathrm{R}}$. Next, we note that

$$\begin{aligned}
\mathbf{A}_{M_\mathrm{R}} &= \mathbf{I}_{M_\mathrm{T}} - \mathbf{G}\mathbf{G}^T \\
&= [\,\mathbf{G}\;\; \mathbf{K}\,] \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_\mathrm{T}-M_\mathrm{R}+1} \end{bmatrix} [\,\mathbf{G}\;\; \mathbf{K}\,]^T \\
&= \sum_{n=M_\mathrm{R}}^{M_\mathrm{T}} \mathbf{u}_n \mathbf{u}_n^T
\end{aligned}$$

with

$$\mathbf{G} = \left[\frac{\mathbf{s}_1^\perp}{\|\mathbf{s}_1^\perp\|} \;\; \frac{\mathbf{s}_2^\perp}{\|\mathbf{s}_2^\perp\|} \;\; \cdots \;\; \frac{\mathbf{s}_{M_\mathrm{R}-1}^\perp}{\|\mathbf{s}_{M_\mathrm{R}-1}^\perp\|} \right] \quad \text{and}$$
$$\mathbf{K} = [\,\mathbf{u}_{M_\mathrm{R}} \;\; \mathbf{u}_{M_\mathrm{R}+1} \;\; \cdots \;\; \mathbf{u}_{M_\mathrm{T}}\,]$$

and the vectors $\mathbf{u}_n$ $(n = M_\mathrm{R}, M_\mathrm{R}+1, \ldots, M_\mathrm{T})$ have to be chosen such that the matrix $\mathbf{U} = [\,\mathbf{G}\;\; \mathbf{K}\,]$ satisfies $\mathbf{U}\mathbf{U}^T = \mathbf{G}\mathbf{G}^T + \mathbf{K}\mathbf{K}^T = \mathbf{I}$. Recognizing that the vectors $\mathbf{s}_i^\perp / \|\mathbf{s}_i^\perp\|$ $(i = 1, 2, \ldots, M_\mathrm{R} - 1)$ are obtained by applying the Gram-Schmidt procedure to the real-valued $(M_\mathrm{R} - 1) \times M_\mathrm{T}$ i.i.d. Gaussian matrix $[\,\mathbf{s}_1 \;\; \mathbf{s}_2 \;\; \cdots \;\; \mathbf{s}_{M_\mathrm{R}-1}\,]^T$ with zero-mean entries, we can take the stacked matrix $\mathbf{U} = [\,\mathbf{G}\;\; \mathbf{K}\,]$ to be given by the Q-matrix obtained by applying the QR-decomposition to an

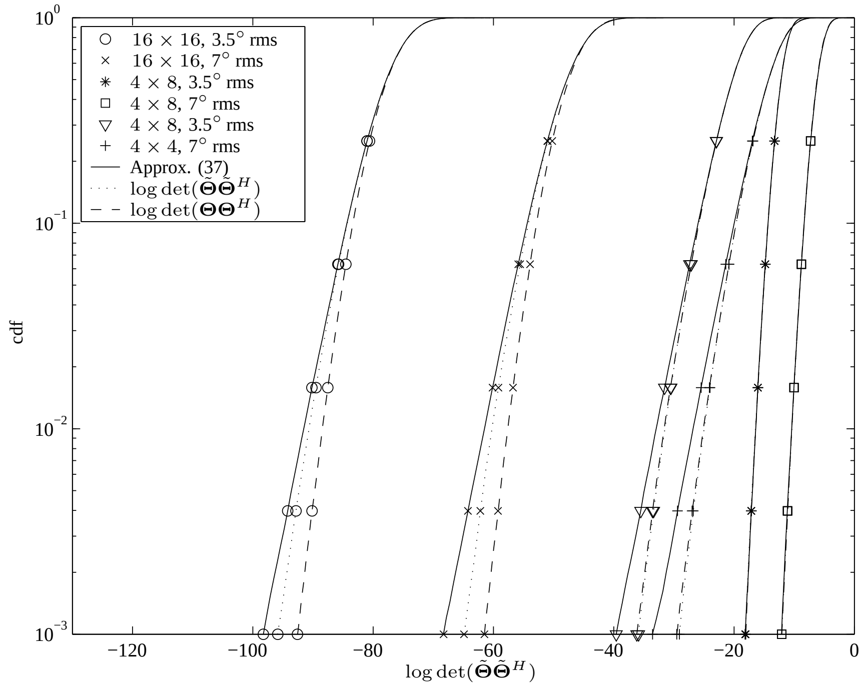

Fig. 2. cdf of $\log\det(\tilde{\mathbf{\Theta}}\tilde{\mathbf{\Theta}}^H)$ and analytic approximation (36), for various $M_{\mathrm{T}}, M_{\mathrm{R}}$ and different phase noise levels. For comparison, $\log\det(\mathbf{\Theta}\mathbf{\Theta}^H)$ is also shown. A regular sounding pattern according to (9) was assumed. All results are based on 300 000 Monte Carlo runs.

$M_{\mathrm{T}} \times M_{\mathrm{T}}$ i.i.d. real-valued Gaussian matrix with zero-mean entries. Note that using the Gram-Schmidt procedure for QR-decomposition yields the unique factorization characterized by positive entries on the main diagonal of the R-matrix [18, Th. 2.6.1]. Next, realizing that

$$\|\mathbf{A}_{M_{\mathrm{R}}}\mathrm{Im}(\mathbf{s}_{M_{\mathrm{R}}})\|^2 = \left(\mathrm{Im}(\mathbf{s}_{M_{\mathrm{R}}})\right)^T \left(\sum_{n=M_{\mathrm{R}}}^{M_{\mathrm{T}}} \mathbf{u}_n\mathbf{u}_n^T\right) \mathrm{Im}(\mathbf{s}_{M_{\mathrm{R}}})$$
$$= M_{\mathrm{T}}M_{\mathrm{R}} \sum_{n=M_{\mathrm{R}}}^{M_{\mathrm{T}}} [\mathbf{U}]^2_{M_{\mathrm{R}},n}$$

the proof is complete upon deriving the pdf of $\sum_{n=M_{\mathrm{R}}}^{M_{\mathrm{T}}}[\mathbf{U}]^2_{M_{\mathrm{R}},n}$. It is well known that, applying any procedure for QR-decomposition leading to the unique factorization where the elements on the main diagonal of the R-matrix are positive, the resulting Q-matrix $\mathbf{Q}$ is distributed such that $\mathbf{AQB} \overset{\mathrm{d}}{=} \mathbf{Q}$ for any orthonormal[8] $\mathbf{A}$ and $\mathbf{B}$ [29, Th. 3.2]. Choosing $\mathbf{A}$ and $\mathbf{B}$ to be permutation matrices, we can conclude that the rows and columns of $\mathbf{Q}$, and hence $\mathbf{U}$ in our case, are all equally distributed. Now, the quantity we are interested in is the sum of squares of the elements $\{M_{\mathrm{R}}, M_{\mathrm{R}}+1, \ldots, M_{\mathrm{T}}\}$ in any such row or column. Specifically, if the Gram-Schmidt procedure is used to obtain the QR-decomposition, the first column of $\mathbf{U}$ is given explicitly as $\mathbf{s}_1/\|\mathbf{s}_1\|$. From [10, Def. 1.4] we know that the quantities $[\mathbf{s}_1]_n^2/\|\mathbf{s}_1\|^2$ are jointly Dirichlet distributed, i.e.,

$$\mathbf{s} = \left[\frac{[\mathbf{s}_1]_1^2}{\|\mathbf{s}_1\|^2} \;\; \frac{[\mathbf{s}_1]_2^2}{\|\mathbf{s}_1\|^2} \;\; \cdots \;\; \frac{[\mathbf{s}_1]_{M_{\mathrm{T}}}^2}{\|\mathbf{s}_1\|^2}\right] \overset{\mathrm{d}}{=} D_{M_{\mathrm{T}}}\left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\right).$$

Partitioning $\mathbf{s}$ into subvectors of length $M_{\mathrm{T}} - M_{\mathrm{R}} + 1$ and $M_{\mathrm{R}} - 1$, respectively, and employing [10, Th. 1.4 and Th. 1.5] (reproduced as Theorem 10 and Theorem 11, respectively, in the Appendix for convenience), it follows that

$$M_{\mathrm{T}}M_{\mathrm{R}} \sum_{n=M_{\mathrm{R}}}^{M_{\mathrm{T}}} [\mathbf{U}]^2_{M_{\mathrm{R}},n}$$
$$\overset{\mathrm{d}}{=} M_{\mathrm{T}}M_{\mathrm{R}}\,\beta\left(\frac{M_{\mathrm{T}} - M_{\mathrm{R}} + 1}{2}, \frac{M_{\mathrm{R}} - 1}{2}\right)$$

where $\beta(a, b)$ is a beta-distributed RV with parameters a and b as defined in the Notations section. ■

Note that even though the results in (36) and Theorem 6 have a striking similarity and (36) provides an approximation for the exact result in Theorem 6, it seems difficult to derive (36) directly from Theorem 6.

We are now ready to state an analytic lower bound on the mean MI of an effective channel resulting from a rank-1 physical channel with unit-modulus entries subject to fully uncorrelated phase noise.

Theorem 7: Under the small phase noise approximation $\sigma_\varphi^2 \ll 1$ so that $\widetilde{\mathbf{\Theta}} = \mathbf{1} + j\mathbf{\Phi}$ with $\mathrm{vec}(\mathbf{\Phi}) \overset{\mathrm{d}}{=} \mathcal{N}(\mathbf{0}, \sigma_\varphi^2\mathbf{I}_{M_{\mathrm{T}}M_{\mathrm{R}}})$, assuming that $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ with $|[\mathbf{g}]_i| = 1$ $(i = 1, 2, \ldots, M_{\mathrm{R}})$ and $|[\mathbf{h}]_i| = 1$ $(i = 1, 2, \ldots, M_{\mathrm{T}})$, the mean MI of the effective channel $\widehat{\mathbf{H}}$ satisfies

$$\widehat{C} \geq \log\left(1 + \sum_{n=1}^{M_{\mathrm{R}}} \left(\frac{\rho}{M_{\mathrm{T}}}\right)^n \binom{M_{\mathrm{R}}}{n} \cdot \prod_{i=0}^{n-1}\left(\delta_i n M_{\mathrm{T}} + 2\sigma_\varphi^2 e^{\Psi\left(\frac{M_{\mathrm{T}}-i}{2}\right)}\right)\right), \quad M_{\mathrm{R}} \leq M_{\mathrm{T}} \tag{41a}$$

[8] The matrix $\mathbf{A}$ is said to be orthonormal if $\mathbf{A}\mathbf{A}^T = \mathbf{I}$.

$$\widehat{C} \geq \log\Bigg(1+\sum_{n=1}^{M_{\mathrm{T}}}\left(\frac{\rho}{M_{\mathrm{T}}}\right)^{n}\binom{M_{\mathrm{T}}}{n} \cdot \prod_{i=0}^{n-1}\left(\delta_i n M_{\mathrm{R}} + 2\sigma_{\varphi}^2 e^{\Psi\left(\frac{M_{\mathrm{R}}-i}{2}\right)}\right)\Bigg), \quad M_{\mathrm{R}} > M_{\mathrm{T}}. \tag{41b}$$

Proof: We provide the proof for $M_{\mathrm{R}} \leq M_{\mathrm{T}}$ only. The case $M_{\mathrm{R}} > M_{\mathrm{T}}$ follows exactly the same line of reasoning. We start by using Lemma 1 and noting that our assumptions imply that

$$\begin{aligned}\hat{I} &= \log\det\left(\mathbf{I} + \frac{\rho}{M_{\mathrm{T}}}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H\right)\\ &= \log\det\left(\mathbf{I} + \frac{\rho}{M_{\mathrm{T}}}\widetilde{\mathbf{\Theta}}\widetilde{\mathbf{\Theta}}^H\right)\\ &= \log\det\left(\mathbf{I} + \frac{\rho}{M_{\mathrm{T}}}\mathbf{S}\mathbf{S}^H\right)\end{aligned} \tag{42}$$

where $\mathbf{S} = -j\mathbf{\Sigma} + \widetilde{\mathbf{\Phi}}$ was defined in the proof of Theorem 5. Next, using [30, Eq. (25)], it follows that

$$\widehat{C} \geq \log\Bigg(1+\sum_{i=1}^{M_{\mathrm{R}}}\left(\frac{\rho}{M_{\mathrm{T}}}\right)^{i} \cdot \sum_{l_1<l_2<\ldots<l_i} e^{\mathbb{E}\left\{\log_e\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\}}\Bigg) \tag{43}$$

where $(\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}$ denotes the submatrix of $\mathbf{S}\mathbf{S}^H$ obtained by retaining the rows $l_1 < l_2 < \ldots < l_i$ and the columns $l_1 < l_2 < \ldots < l_i$. The summation in (43) is over all ordered tuples $(l_1, l_2, \ldots, l_i)$ chosen from the set $\{1, 2, \ldots, M_{\mathrm{R}}\}$. Next, we note that the pdf of $\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)$ follows in a straightforward fashion from the results developed in the proof of Theorem 5. In particular, $\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)$ is the determinant of the matrix $\widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^H$ where the $i \times M_{\mathrm{T}}$ matrix $\widetilde{\mathbf{S}}$ is obtained from $\mathbf{S}$ by retaining the rows $\{l_1, l_2, \ldots, l_i\}$. Distinguishing between the terms, in the summation over $l_1 < l_2 < \ldots < l_i$ on the RHS of (43), that have $l_1 = 1$ and those where $l_1 > 1$, we obtain

$$\mathbb{E}\left\{\log_e\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\} \overset{(a)}{\geq} \mathbb{E}\left\{\log_e\left(\left(\chi^2_{M_{\mathrm{T}},\sigma_\varphi^2} + M_{\mathrm{T}}M_{\mathrm{R}}\right)\prod_{l=1}^{i-1}\chi^2_{M_{\mathrm{T}}-l,\sigma_\varphi^2}\right)\right\} \tag{44}$$

in the former case and

$$\mathbb{E}\left\{\log_e\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\} = \mathbb{E}\left\{\log_e\left(\prod_{l=0}^{i-1}\chi^2_{M_{\mathrm{T}}-l,\sigma_\varphi^2}\right)\right\} \tag{45}$$

in the latter case, where (a) is obtained as follows. Recognizing that $\det((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}) = \det((\widetilde{\mathbf{\Theta}}\widetilde{\mathbf{\Theta}}^H)_{l_1<l_2<\ldots<l_i})$ and applying (36) properly modified to account for the fact that we are interested in the submatrix of $\widetilde{\mathbf{\Theta}}\widetilde{\mathbf{\Theta}}^H$ obtained by retaining the rows $\{l_1 < l_2 < \ldots < l_i\}$ in $\widetilde{\mathbf{\Theta}}$ would yield an approximate expression for the left-hand side (LHS) in (44). However, using Theorem 5 and invoking Theorem 12 in the Appendix, we can show that the lower bound in (44) holds firmly. Specifically, starting from (30) and setting

$$\mathfrak{X}'_{M_{\mathrm{T}},\sigma_\varphi^2} = \mathbb{E}\{\log_e(M_{\mathrm{T}}M_{\mathrm{R}} + \chi^2_{M_{\mathrm{T}},\sigma_\varphi^2})\},$$

for brevity, we can rewrite the LHS of (44) as

$$\begin{aligned}&\mathbb{E}\left\{\log_e\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\}\\ &= \mathfrak{X}'_{M_{\mathrm{T}},\sigma_\varphi^2} + \sum_{l=2}^{i}\mathbb{E}_{X_l,X_1,X_2,\eta^{(l)}}\Big\{\log_e\big(\underbrace{\chi^2_{M_{\mathrm{T}}-l,\sigma_\varphi^2}}_{X_l} + Z(\eta^{(l)})\big)\Big\}\\ &= \mathfrak{X}'_{M_{\mathrm{T}},\sigma_\varphi^2} + \sum_{l=2}^{i}\mathbb{E}_{X_l}\mathbb{E}_{\eta^{(l)}}\mathbb{E}_{X_1,X_2|X_l,\eta^{(l)}}\left\{\log_e\left(X_l + Z(\eta^{(l)})\right)\right\}\\ &\overset{(a)}{\geq} \mathfrak{X}'_{M_{\mathrm{T}},\sigma_\varphi^2} + \sum_{l=2}^{i}\mathbb{E}_{X_l}\mathbb{E}_{Y|X_l}\Big\{\log_e\big(X_l + \underbrace{\chi^2_{1,\sigma_\varphi^2}}_{Y}\big)\Big\}\\ &= \mathfrak{X}'_{M_{\mathrm{T}},\sigma_\varphi^2} + \sum_{l=1}^{i-1}\mathbb{E}\left\{\log_e\left(\chi^2_{M_{\mathrm{T}}-l,\sigma_\varphi^2}\right)\right\}\end{aligned} \tag{46}$$

where (a) follows from Theorem 12 in the Appendix. The relation in (45) is obtained in exactly the same fashion upon noting that the term $-j\sqrt{M_{\mathrm{T}}M_{\mathrm{R}}}$ is absent in the sets $(l_1, l_2, \ldots, l_i)$ where $l_1 > 1$. The number of terms in the first group (where $l_1 = 1$) is given by $\binom{M_{\mathrm{R}}-1}{i-1}$ whereas the number of terms in the second group is $\binom{M_{\mathrm{R}}}{i} - \binom{M_{\mathrm{R}}-1}{i-1}$. It remains to find analytic expressions for the RHS of (44) and of (45).

It is well known [31] that

$$\mathbb{E}\{\log_e(\chi^2_{n,\sigma_\varphi^2})\} = \log_e(2\sigma_\varphi^2) + \Psi(n/2) \triangleq \mathfrak{X}_{n,\sigma_\varphi^2}.$$

The term $\mathfrak{X}'_{M_{\mathrm{T}},\sigma_\varphi^2} = \mathbb{E}\{\log_e(\chi^2_{M_{\mathrm{T}},\sigma_\varphi^2} + M_{\mathrm{T}}M_{\mathrm{R}})\}$ has a closed-form analytic expression in terms of the generalized exponential integral

$$E_\nu(z) = \int_1^\infty t^{-\nu}e^{-zt}\,\mathrm{d}t, \quad \mathrm{Re}(z) > 0.$$

For our purposes, we shall, however, be content with a simple lower bound obtained by applying Jensen's inequality to the function $f(x) = \log_e(e^x + a)$, which results in

$$\begin{aligned}\mathfrak{X}'_{M_{\mathrm{T}},\sigma_\varphi^2} &\geq \log_e\left(e^{\mathbb{E}\left\{\log_e\left(\chi^2_{M_{\mathrm{T}},\sigma_\varphi^2}\right)\right\}} + M_{\mathrm{T}}M_{\mathrm{R}}\right)\\ &= \log_e\left(e^{\mathfrak{X}_{M_{\mathrm{T}},\sigma_\varphi^2}} + M_{\mathrm{T}}M_{\mathrm{R}}\right).\end{aligned} \tag{47}$$

Putting the pieces together, we get

$$\begin{aligned}&e^{\mathbb{E}\left\{\log_e\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\}}\\ &\geq \left(M_{\mathrm{T}}M_{\mathrm{R}} + e^{\mathfrak{X}_{M_{\mathrm{T}},\sigma_\varphi^2}}\right)e^{\sum_{l=1}^{i-1}\mathfrak{X}_{M_{\mathrm{T}}-l,\sigma_\varphi^2}}\\ &= M_{\mathrm{T}}M_{\mathrm{R}}e^{\sum_{l=1}^{i-1}\mathfrak{X}_{M_{\mathrm{T}}-l,\sigma_\varphi^2}} + e^{\sum_{l=0}^{i-1}\mathfrak{X}_{M_{\mathrm{T}}-l,\sigma_\varphi^2}}\end{aligned}$$

for (44) and

$$e^{\mathbb{E}\left\{\log_e\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\}} = e^{\sum_{l=0}^{i-1}\mathfrak{X}_{M_{\mathrm{T}}-l,\sigma_\varphi^2}}$$

for (45). Combining our results and noting that the term

$\exp\big(\sum_{l=0}^{i-1}\mathfrak{X}_{M_\mathrm{T}-l,\sigma_\varphi^2}\big)$ occurs in both cases so that its total number of occurences is $\binom{M_\mathrm{R}}{i}$, finally yields

$$\sum_{l_1<l_2<\ldots<l_i} e^{\mathbb{E}\left\{\log_e \det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\}} \geq \binom{M_\mathrm{R}-1}{i-1} M_\mathrm{T} M_\mathrm{R} e^{\sum_{n=1}^{i-1}\mathfrak{X}_{M_\mathrm{T}-n,\sigma_\varphi^2}} + \binom{M_\mathrm{R}}{i} e^{\sum_{n=0}^{i-1}\mathfrak{X}_{M_\mathrm{T}-n,\sigma_\varphi^2}} \tag{48}$$

which, upon inserting into (43) and reorganizing terms, concludes the proof. ■

The result in (41) can be made more explicit by using the simplifications for the digamma function at positive integer multiples of $1/2$ given by (3). Furthermore, we note that Theorem 7 can be generalized to the cases where i) $|[\mathbf{h}]_i|=1$, $\forall\, i$, $\mathbf{g}$ is general and $M_\mathrm{R}\leq M_\mathrm{T}$ and ii) $|[\mathbf{g}]_i|=1$, $\forall\, i$, $\mathbf{h}$ is general and $M_\mathrm{R}>M_\mathrm{T}$. The corresponding results are stated, without proof, as

$$\text{i)}\quad \widehat{C} \geq \log\left(1+\sum_{n=1}^{M_\mathrm{R}}\left(\frac{\rho}{M_\mathrm{T}}\right)^n K_n(\mathbf{g}) \cdot \prod_{i=0}^{n-1}\left(\delta_i n M_\mathrm{T} + 2\sigma_\varphi^2 e^{\Psi\left(\frac{M_\mathrm{T}-i}{2}\right)}\right)\right), \quad M_\mathrm{R}\leq M_\mathrm{T}$$

$$\text{ii)}\quad \widehat{C} \geq \log\left(1+\sum_{n=1}^{M_\mathrm{T}}\left(\frac{\rho}{M_\mathrm{T}}\right)^n K_n(\mathbf{h}) \cdot \prod_{i=0}^{n-1}\left(\delta_i n M_\mathrm{R} + 2\sigma_\varphi^2 e^{\Psi\left(\frac{M_\mathrm{R}-i}{2}\right)}\right)\right), \quad M_\mathrm{R}> M_\mathrm{T}$$

with

$$K_n(\mathbf{x}) = \sum_{\mathbf{s}\in\mathcal{S}_{n,l(\mathbf{x})}} \prod_{i=1}^{n} |[\mathbf{x}]_{s_i}|^2,$$

where $\mathcal{S}_{k,m}$ is the set of all possible ordered k-tuples $\mathbf{s}=(s_1,s_2,\ldots,s_k)$ with $1\leq s_1<s_2<\ldots<s_k\leq m$, and $l(\mathbf{x})$ is the number of elements in the vector $\mathbf{x}$.

Again assuming $|[\mathbf{g}]_i|=1$ $(i=1,2,\ldots,M_\mathrm{R})$ and $|[\mathbf{h}]_i|=1$ $(i=1,2,\ldots,M_\mathrm{T})$, further lower-bounding (41) by ignoring the first term inside the "log" and retaining only the highest-order (in ρ) term yields

$$\widehat{C} \geq M_\mathrm{R}\log\left(\frac{\rho}{M_\mathrm{T}}\right) + \log\left(\prod_{i=0}^{M_\mathrm{R}-1}\left(\delta_i M_\mathrm{T}M_\mathrm{R} + 2\sigma_\varphi^2 e^{\Psi\left(\frac{M_\mathrm{T}-i}{2}\right)}\right)\right), \quad M_\mathrm{R}\leq M_\mathrm{T}$$

$$\widehat{C} \geq M_\mathrm{T}\log\left(\frac{\rho}{M_\mathrm{T}}\right) + \log\left(\prod_{i=0}^{M_\mathrm{T}-1}\left(\delta_i M_\mathrm{T}M_\mathrm{R} + 2\sigma_\varphi^2 e^{\Psi\left(\frac{M_\mathrm{R}-i}{2}\right)}\right)\right), \quad M_\mathrm{R}> M_\mathrm{T}$$

which clearly shows that the effective channel has full rank and hence its multiplexing gain is given by $\min(M_\mathrm{T},M_\mathrm{R})$. Put differently, phase noise can cause a rank-1 physical channel to appear like a full-rank channel. At the end of this section, we shall show, based on measurement results, that significant rank increase does, indeed, occur in practice.

We shall next provide a slightly looser (than (41)) lower bound on $\widehat{C}$ with a simpler structure.

Theorem 8: Under the small phase noise approximation $\sigma_\varphi^2 \ll 1$ so that $\widetilde{\mathbf{\Theta}} = \mathbf{1}+j\mathbf{\Phi}$ with $\mathrm{vec}(\mathbf{\Phi}) \overset{\mathrm{d}}{=} \mathcal{N}(\mathbf{0},\sigma_\varphi^2\mathbf{I}_{M_\mathrm{T}M_\mathrm{R}})$, assuming that $\mathbf{H}=\mathbf{g}\mathbf{h}^T$ with $|[\mathbf{g}]_i|=1$ $(i=1,2,\ldots,M_\mathrm{R})$ and $|[\mathbf{h}]_i|=1$ $(i=1,2,\ldots,M_\mathrm{T})$, the mean MI of the effective channel $\widehat{\mathbf{H}}$ satisfies

$$\widehat{C} \geq \sum_{i=0}^{M_\mathrm{R}-1}\log\left(1+\frac{\rho}{M_\mathrm{T}}\left(M_\mathrm{T}M_\mathrm{R}\delta_i + 2\sigma_\varphi^2 e^{\Psi\left(\frac{M_\mathrm{T}-i}{2}\right)}\right)\right), \quad M_\mathrm{R}\leq M_\mathrm{T} \tag{49a}$$

$$\widehat{C} \geq \sum_{i=0}^{M_\mathrm{T}-1}\log\left(1+\frac{\rho}{M_\mathrm{T}}\left(M_\mathrm{T}M_\mathrm{R}\delta_i + 2\sigma_\varphi^2 e^{\Psi\left(\frac{M_\mathrm{R}-i}{2}\right)}\right)\right), \quad M_\mathrm{R}> M_\mathrm{T}. \tag{49b}$$

Proof: We provide the proof for $M_\mathrm{R}\leq M_\mathrm{T}$ only. The case $M_\mathrm{R}>M_\mathrm{T}$ follows exactly the same line of reasoning. We start by noting that (44) and (45) can be combined as

$$\mathbb{E}\left\{\log_e\det\left((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_i}\right)\right\} \geq \sum_{l=0}^{i-1}\mathbb{E}\left\{\log_e\left(M_\mathrm{T}M_\mathrm{R}\delta_l\delta_{l_1-1}+\chi^2_{M_\mathrm{T}-l,\sigma_\varphi^2}\right)\right\}$$

which, upon inserting into (43), yields

$$\widehat{C} \geq \log\left(1+\sum_{i=1}^{M_\mathrm{R}}\left(\frac{\rho}{M_\mathrm{T}}\right)^i \cdot \sum_{l_1<l_2<\ldots<l_i}\prod_{l=0}^{i-1} e^{\mathbb{E}\left\{\log_e\left(M_\mathrm{T}M_\mathrm{R}\delta_l\delta_{l_1-1}+\chi^2_{M_\mathrm{T}-l,\sigma_\varphi^2}\right)\right\}}\right) = C_2.$$

The proof will be completed by showing that $2^{C_2}\geq 2^{C_1}$ with

$$C_1 = \log\prod_{i=0}^{M_\mathrm{R}-1}\left(1+\frac{\rho}{M_\mathrm{T}}e^{\mathbb{E}\left\{\log_e\left(M_\mathrm{T}M_\mathrm{R}\delta_i+\chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2}\right)\right\}}\right) \tag{50}$$

and noting that the RHS of (49) is obtained by lower-bounding the term corresponding to $i=0$ in (50) according to (47). Setting

$$X_n = \mathfrak{X}_{M_\mathrm{T}-n,\sigma_\varphi^2}, \quad n=0,1,\ldots,M_\mathrm{R}-1$$
$$X_0' = \mathbb{E}\{\log_e(M_\mathrm{T}M_\mathrm{R}+\chi^2_{M_\mathrm{T},\sigma_\varphi^2})\},$$

and expanding

$$2^{C_1} = \left(1+\frac{\rho}{M_\mathrm{T}}e^{X_0'}\right)\prod_{i=1}^{M_\mathrm{R}-1}\left(1+\frac{\rho}{M_\mathrm{T}}e^{X_i}\right)$$

we get

$$\begin{aligned}
&2^{C_1} - 1 \\
&\quad= \left(\frac{\rho}{M_{\mathrm{T}}}\right)\left(e^{X_0'} + e^{X_1} + \cdots + e^{X_{M_{\mathrm{R}}-1}}\right) \\
&\text{2a)} \quad + \left(\frac{\rho}{M_{\mathrm{T}}}\right)^2 \Big(e^{X_0'}e^{X_1} + e^{X_0'}e^{X_2} + \cdots \\
&\text{2b)} \qquad\qquad\qquad + e^{X_0'}e^{X_{M_{\mathrm{R}}-1}} \\
&\text{2c)} \qquad + e^{X_1}e^{X_2} + e^{X_1}e^{X_3} + \cdots \\
&\text{2d)} \qquad + \cdots + e^{X_{M_{\mathrm{R}}-2}}e^{X_{M_{\mathrm{R}}-1}}\Big) \\
&\text{3a)} \quad + \left(\frac{\rho}{M_{\mathrm{T}}}\right)^3 \Big(e^{X_0'}e^{X_1}e^{X_2} + e^{X_0'}e^{X_1}e^{X_3} + \cdots \\
&\text{3b)} \qquad\qquad + e^{X_0'}e^{X_{M_{\mathrm{R}}-2}}e^{X_{M_{\mathrm{R}}-1}} \\
&\text{3c)} \qquad + e^{X_1}e^{X_2}e^{X_3} + \cdots \\
&\text{3d)} \qquad + \ldots + e^{X_{M_{\mathrm{R}}-3}}e^{X_{M_{\mathrm{R}}-2}}e^{X_{M_{\mathrm{R}}-1}}\Big) \\
&\quad + \quad \vdots \qquad\qquad \vdots \\
&\quad + \left(\frac{\rho}{M_{\mathrm{T}}}\right)^{M_{\mathrm{R}}} \left(e^{X_0'}e^{X_1}\cdots e^{X_{M_{\mathrm{R}}-1}}\right)
\end{aligned} \tag{51}$$

where lines 2a)-2b) and 2c)-2d) contain $\binom{M_{\mathrm{R}}-1}{1}$ and $\binom{M_{\mathrm{R}}-1}{2}$ terms, respectively, and lines 3a)-3b) and 3c)-3d) contain $\binom{M_{\mathrm{R}}-1}{2}$ and $\binom{M_{\mathrm{R}}-1}{3}$ terms, respectively. Expanding

$$\begin{aligned}
2^{C_2} = 1 + \sum_{i=1}^{M_{\mathrm{R}}} \left(\frac{\rho}{M_{\mathrm{T}}}\right)^i
&\cdot \left(\sum_{1=l_1<l_2<\ldots<l_i} \left(X_0' \prod_{l=1}^{i-1} e^{X_l} \right) \right. \\
&\left. + \sum_{1<l_1<l_2<\ldots<l_i} \left(X_0 \prod_{l=1}^{i-1} e^{X_l} \right) \right)
\end{aligned}$$

we obtain

$$\begin{aligned}
&2^{C_2} - 1 \\
&\quad= \left(\frac{\rho}{M_{\mathrm{T}}}\right)\left(e^{X_0'} + e^{X_0} + \cdots + e^{X_0}\right) \\
&\text{2a)} \quad + \left(\frac{\rho}{M_{\mathrm{T}}}\right)^2 \Big(e^{X_0'}e^{X_1} + e^{X_0'}e^{X_1} + \cdots \\
&\text{2b)} \qquad\qquad\qquad + e^{X_0'}e^{X_1} \\
&\text{2c)} \qquad + e^{X_0}e^{X_1} + e^{X_0}e^{X_1} + \cdots \\
&\text{2d)} \qquad + \cdots + e^{X_0}e^{X_1}\Big) \\
&\text{3a)} \quad + \left(\frac{\rho}{M_{\mathrm{T}}}\right)^3 \Big(e^{X_0'}e^{X_1}e^{X_2} + e^{X_0'}e^{X_1}e^{X_2} + \cdots \\
&\text{3b)} \qquad\qquad\qquad + e^{X_0'}e^{X_1}e^{X_2} \\
&\text{3c)} \qquad + e^{X_0}e^{X_1}e^{X_2} + e^{X_0}e^{X_1}e^{X_2} + \cdots \\
&\text{3d)} \qquad + \cdots + e^{X_0}e^{X_1}e^{X_2}\Big) \\
&\quad + \quad \vdots \qquad\qquad \vdots \\
&\quad + \left(\frac{\rho}{M_{\mathrm{T}}}\right)^{M_{\mathrm{R}}} \left(e^{X_0'}e^{X_1}\cdots e^{X_{M_{\mathrm{R}}-1}}\right)
\end{aligned} \tag{52}$$

where the number of terms in lines 2a)-2b), 2c)-2d), 3a)-3b), and 3c)-3d) is the same as in the corresponding lines in the expansion of 2^{C_1}. In both cases, the number of terms associated with the factor $(\rho/M_{\mathrm{T}})^i$ is given by $\binom{M_{\mathrm{R}}}{i}$. The proof is completed by comparing (51) and (52) term by term and noting that the monotonicity of the digamma function implies $\mathfrak{X}_{M_{\mathrm{T}}-n,\sigma_\varphi^2} \geq \mathfrak{X}_{M_{\mathrm{T}}-n-k,\sigma_\varphi^2}$ and hence $X_n \geq X_{n+k}$ for $k \geq 1$. ■

So far we derived lower bounds on $\widehat{C}$. We shall next show that the result in Theorem 5 together with a technique first proposed in [30] (and used to derive the lower bound in Theorem 7) can be employed to derive a tight analytic upper bound on $\widehat{C}$.

Theorem 9: Under the assumptions in Theorem 8, the mean MI of the effective MIMO channel can be upper-bounded as

$$\begin{aligned}
\widehat{C} \leq \log\Bigg(1 + \sum_{n=1}^{M_{\mathrm{R}}} \left(\frac{\rho}{M_{\mathrm{T}}}\right)^n \\
\cdot (\sigma_\varphi^2)^n \binom{M_{\mathrm{R}}}{n}\binom{M_{\mathrm{T}}}{n} n! \left(1 + \frac{n}{\sigma_\varphi^2}\right)\Bigg).
\end{aligned} \tag{53}$$

Proof: We provide the proof for $M_{\mathrm{R}} \leq M_{\mathrm{T}}$ only. The case $M_{\mathrm{R}} > M_{\mathrm{T}}$ follows exactly the same line of reasoning. The proof starts from [30, Eq. (19)] which, specialized to our case, reads

$$\begin{aligned}
\widehat{C} \leq \log\Bigg(1 + \sum_{n=1}^{M_{\mathrm{R}}} \left(\frac{\rho}{M_{\mathrm{T}}}\right)^n \\
\cdot \sum_{l_1<l_2<\ldots<l_n} \mathbb{E}\big\{\det\big((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_n}\big)\big\}\Bigg)
\end{aligned} \tag{54}$$

where $\mathbf{S}$ was defined in the proof of Theorem 5. The main point of the proof is to recognize that we can obtain analytic expressions for the terms $\mathbb{E}\{\det((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_n})\}$ using (32). As already shown in the proof of Theorem 7, the terms in $\sum_{l_1<l_2<\ldots<l_n} \mathbb{E}\{\cdot\}$ on the RHS of (54) fall into two groups depending on whether $l_1 = 1$ or $l_1 > 1$. Specifically, for $l_1 = 1$ we have (cf. (46))

$$\begin{aligned}
&\mathbb{E}\{\det((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_n})\} \\
&\quad= M_{\mathrm{T}}M_{\mathrm{R}} \prod_{i=2}^{n} \mathbb{E}\big\{\chi^2_{M_{\mathrm{T}}-i,\sigma_\varphi^2} + Z(\eta^{(i)})\big\} \\
&\qquad + \mathbb{E}\big\{\chi^2_{M_{\mathrm{T}},\sigma_\varphi^2}\big\} \prod_{i=2}^{n} \mathbb{E}\big\{\chi^2_{M_{\mathrm{T}}-i,\sigma_\varphi^2} + Z(\eta^{(i)})\big\}
\end{aligned}$$

and for $l_1 > 1$ (cf. (45))

$$\begin{aligned}
&\mathbb{E}\big\{\det\big((\mathbf{S}\mathbf{S}^H)_{l_1<l_2<\ldots<l_n}\big)\big\} \\
&\quad= \mathbb{E}\big\{\chi^2_{M_{\mathrm{T}},\sigma_\varphi^2}\big\} \prod_{i=2}^{n} \mathbb{E}\big\{\chi^2_{M_{\mathrm{T}}-i,\sigma_\varphi^2} + Z(\eta^{(i)})\big\}.
\end{aligned}$$

Using $\mathbb{E}\{\chi^2_{M_{\mathrm{T}}-i,\sigma_\varphi^2}\} = (M_{\mathrm{T}} - i)\sigma_\varphi^2$, noting that

$$\begin{aligned}
&\mathbb{E}\big\{Z(\eta^{(i)})\big\} \\
&\quad= \mathbb{E}_{\eta^{(i)}}\Big\{\mathbb{E}\big\{\chi^2_{1,\sigma_\varphi^2\eta^{(i)}} + \chi^2_{1,\sigma_\varphi^2(1-\eta^{(i)})} \big| \eta^{(i)}\big\}\Big\} \\
&\quad= \mathbb{E}_{\eta^{(i)}}\big\{\sigma_\varphi^2\eta^{(i)} + \sigma_\varphi^2(1-\eta^{(i)})\big\} \\
&\quad= \mathbb{E}_{\eta^{(i)}}\big\{\sigma_\varphi^2\big\} \\
&\quad= \sigma_\varphi^2
\end{aligned}$$

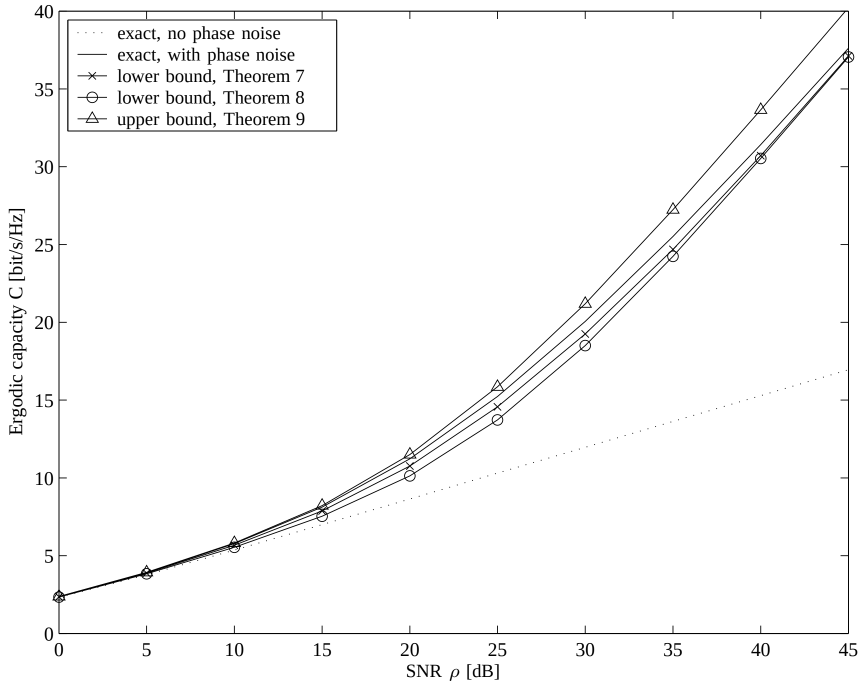

Fig. 3. Mean MI of a 4×4 rank-1 physical channel with unit-modulus entries subject to $7°$ rms fully uncorrelated phase noise. Exact results are obtained through Monte Carlo simulation.

and counting the multiplicity of the terms as in (48), we obtain

$$\begin{aligned}\sum_{l_1<l_2<\ldots<l_n} &\mathbb{E}\{\det((\mathbf{SS}^H)_{l_1<l_2<\ldots<l_n})\}\\ &= \binom{M_\mathrm{R}-1}{n-1} M_\mathrm{T} M_\mathrm{R} \prod_{i=2}^{n} \left((M_\mathrm{T}-i+1)\sigma_\varphi^2\right)\\ &\quad + \binom{M_\mathrm{R}}{n} M_\mathrm{T}\sigma_\varphi^2 \prod_{i=2}^{n} \left((M_\mathrm{T}-i+1)\sigma_\varphi^2\right)\\ &= \binom{M_\mathrm{R}-1}{n-1} M_\mathrm{T} M_\mathrm{R} \frac{(M_\mathrm{T}-1)!}{(M_\mathrm{T}-n)!}(\sigma_\varphi^2)^{n-1}\\ &\quad + \binom{M_\mathrm{R}}{n} M_\mathrm{T} \frac{(M_\mathrm{T}-1)!}{(M_\mathrm{T}-n)!}(\sigma_\varphi^2)^{n}\\ &= \binom{M_\mathrm{R}}{n} \frac{n M_\mathrm{T}!}{(M_\mathrm{T}-n)!}(\sigma_\varphi^2)^{n-1}\\ &\quad + \binom{M_\mathrm{R}}{n} \frac{M_\mathrm{T}!}{(M_\mathrm{T}-n)!}(\sigma_\varphi^2)^{n}.\end{aligned}$$

Putting the pieces together, we get (53), which concludes the proof. ■

We note that the proof of Theorem 9 can alternatively be carried out by obtaining analytic expressions for the terms $\mathbb{E}\{\det((\mathbf{SS}^H)_{l_1<l_2<\ldots<l_n})\}$ using properly modified versions of (39).

Numerical results: We shall next provide a numerical example that serves to quantify the quality of the lower bounds in Theorems 7 and 8, and the upper bound in Theorem 9. For a 4×4 deterministic physical channel $\mathbf{H} = \mathbf{gh}^T$ with $|[\mathbf{g}]_i| = 1$, $\forall i$, and $|[\mathbf{h}]_i| = 1$, $\forall i$, subject to $7°$ rms fully uncorrelated phase noise, Fig. 3 shows that the (ergodic) capacity of the effective channel starts deviating from the capacity of the rank-1 physical MIMO channel at $\rho \approx 15\,\mathrm{dB}$, and that significant capacity estimation errors (up to around 100%) occur in the high-SNR regime. This behavior is consistent with our observation that the low-SNR capacity is not influenced by phase noise. Moreover, we observe that the lower and upper bounds (41) and (53), respectively, very accurately predict the capacity behavior of the effective channel.

Comparison with measurement results: Next, we demonstrate that the capacity bounds in Theorems 7 and 9 match very well with measurement results taken with a commercially employed TDMS-based MIMO channel sounder. The sounder is set up according to the "calibration procedure", discussed in the first paragraph of Section V, where transmitter and receiver are directly connected through a cable. This yields measurements of a rank-1 physical channel and is well suited to illustrate the impact of phase errors on estimated MI. Furthermore, 16 transmit and 16 receive antennas, and a length-511 sounding sequence are used. In Fig. 4, we show the MI obtained from the measurements, averaged over $L = 1100$ MIMO snapshots at each SNR, along with the mean MI predicted by our analytic results in Theorems 7 and 9 for $3.8°$ rms fully uncorrelated phase noise, a level that was estimated from our measurements. We can see that the analytic lower and upper bounds are slightly higher than the measured mean MI. This is probably due to the residual correlation in the phase noise process, neglected in the analytic results. Apart from this effect, the measurement results exhibit an excellent match with the analytic results.

High-SNR Variance of MI: Considering an ergodic block-fading MIMO channel, it was shown in [32] that the (high-SNR) variance of MI can be interpreted as quantifying the amount of "spatial averaging" that occurs *on a per-stream basis* in each fading block. The smaller the variance of MI

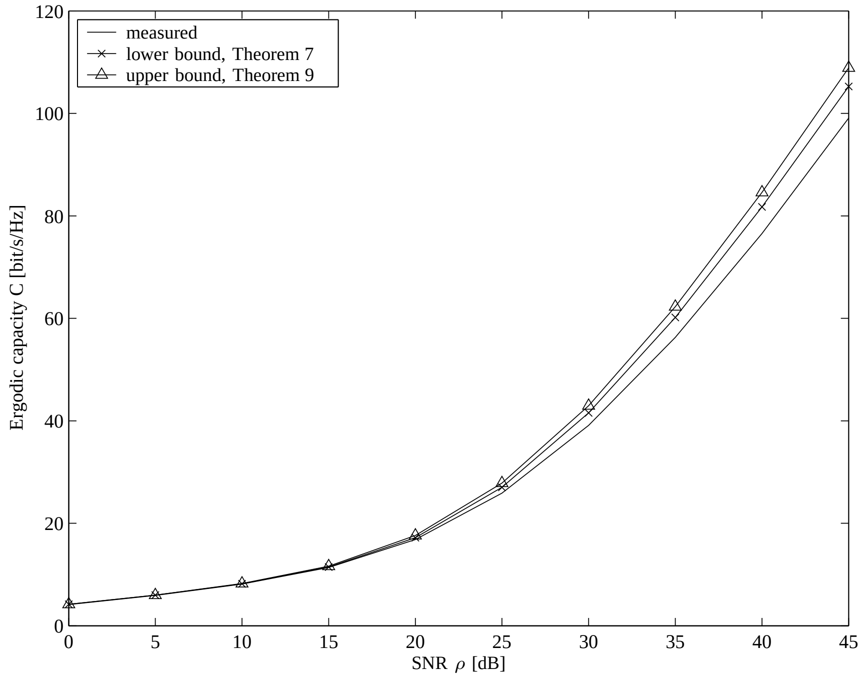

Fig. 4. Measured mean MI along with corresponding analytic lower and upper bound in Theorems 7 and 9, respectively.

the more spatial averaging occurs. As shown in [32], [33], $\sigma_I^2 = \mathrm{Var}\{I\}$ for fixed M_R, as a function of M_T, has its maximum at $M_\mathrm{T} = M_\mathrm{R}$. For more details on the interpretation of σ_I^2 as a measure of the amount of spatial diversity, the interested reader is referred to [32]. We have seen that phase noise (and frequency offset) can have a significant impact on the rank of the MIMO channel and hence its spatial multiplexing gain. In the following, we shall characterize the increase in spatial diversity due to phase noise by analyzing the variance of the high-SNR MI of the effective MIMO channel. Finding exact expressions for $\sigma_{\hat{I}}^2 = \mathrm{Var}\{\hat{I}\}$ seems difficult. Under the assumptions in Theorem 8, we can, however, provide accurate and analytically tractable approximations, which are obtained as follows. Considering, for simplicity, the case $M_\mathrm{R} \le M_\mathrm{T}$, we can infer from (42) and (36) that in the high-SNR regime

$$\begin{aligned}\hat{I} &\approx \log\left(\frac{\rho}{M_\mathrm{T}}\left(\chi^2_{M_\mathrm{T},\sigma_\varphi^2} + M_\mathrm{T}M_\mathrm{R}\right)\prod_{i=1}^{M_\mathrm{R}-1}\chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2}\right)\\ &= \log\left(\frac{\rho}{M_\mathrm{T}}\right) + \log\left(M_\mathrm{T}M_\mathrm{R} + \chi^2_{M_\mathrm{T},\sigma_\varphi^2}\right)\\ &\quad + \sum_{i=1}^{M_\mathrm{R}-1}\log\left(\chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2}\right).\end{aligned} \tag{55}$$

Writing the second term in (55) as

$$\begin{aligned}&\log\left(M_\mathrm{T}M_\mathrm{R} + \chi^2_{M_\mathrm{T},\sigma_\varphi^2}\right)\\ &\quad = \log(M_\mathrm{T}M_\mathrm{R}) + \log\left(1 + \frac{\chi^2_{M_\mathrm{T},\sigma_\varphi^2}}{M_\mathrm{T}M_\mathrm{R}}\right)\end{aligned}$$

and noting that for $M_\mathrm{T}M_\mathrm{R}$ large and σ_φ^2 small, we have

$$\log\left(1 + \frac{\chi^2_{M_\mathrm{T},\sigma_\varphi^2}}{M_\mathrm{T}M_\mathrm{R}}\right) \approx \log(e)\frac{\chi^2_{M_\mathrm{T},\sigma_\varphi^2}}{M_\mathrm{T}M_\mathrm{R}}$$

it follows that

$$\begin{aligned}&\mathrm{Var}\{\hat{I}\}\\ &\quad\approx \mathrm{Var}\left\{\log(e)\frac{\chi^2_{M_\mathrm{T},\sigma_\varphi^2}}{M_\mathrm{T}M_\mathrm{R}} + \sum_{i=1}^{M_\mathrm{R}-1}\log\left(\chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2}\right)\right\}.\end{aligned}$$

Using

$$\begin{aligned}&\mathrm{Var}\left\{\log_e\left(\chi^2_{M_\mathrm{T}-i,\sigma_\varphi^2}\right)\right\}\\ &\quad = \mathrm{Var}\left\{\log_e(\sigma_\varphi^2) + \log_e\left(\chi^2_{M_\mathrm{T}-i,1}\right)\right\}\\ &\quad = \mathrm{Var}\left\{\log_e\left(\chi^2_{M_\mathrm{T}-i,1}\right)\right\}\\ &\quad = \Psi'\left(\frac{M_\mathrm{T}-i}{2}\right)\end{aligned}$$

and [31, App. A.7], we finally get

$$\begin{aligned}&\mathrm{Var}\{\hat{I}\}\\ &\quad\approx \left(\log(e)\right)^2\left(\frac{2\sigma_\varphi^4}{M_\mathrm{T}M_\mathrm{R}^2} + \sum_{i=1}^{M_\mathrm{R}-1}\sum_{p=1}^{\infty}\frac{1}{\left(p + \frac{M_\mathrm{T}-i}{2} - 1\right)^2}\right).\end{aligned} \tag{56}$$

Comparing (56) to the expression [32, Eq. (31)] for the variance of the high-SNR MI of an i.i.d. $M_\mathrm{R} \times M_\mathrm{T}$ complex Gaussian channel, we can show that, for the same number of transmit and receive antennas, the variance of the high-SNR MI of a rank-1 physical channel as in Theorem 8 subject to fully uncorrelated phase noise is higher than that in the i.i.d. complex Gaussian case, i.e.,

$$\frac{2\sigma_\varphi^4}{M_\mathrm{T}M_\mathrm{R}^2} + \sum_{i=1}^{M_\mathrm{R}-1}\Psi'\left(\frac{M_\mathrm{T}-i}{2}\right) \ge \sum_{i=1}^{M_\mathrm{R}}\Psi'(M_\mathrm{T}-i+1). \tag{57}$$

To prove (57), we omit the first term on the LHS and use Lemma 3 in the Appendix which leaves us with having to

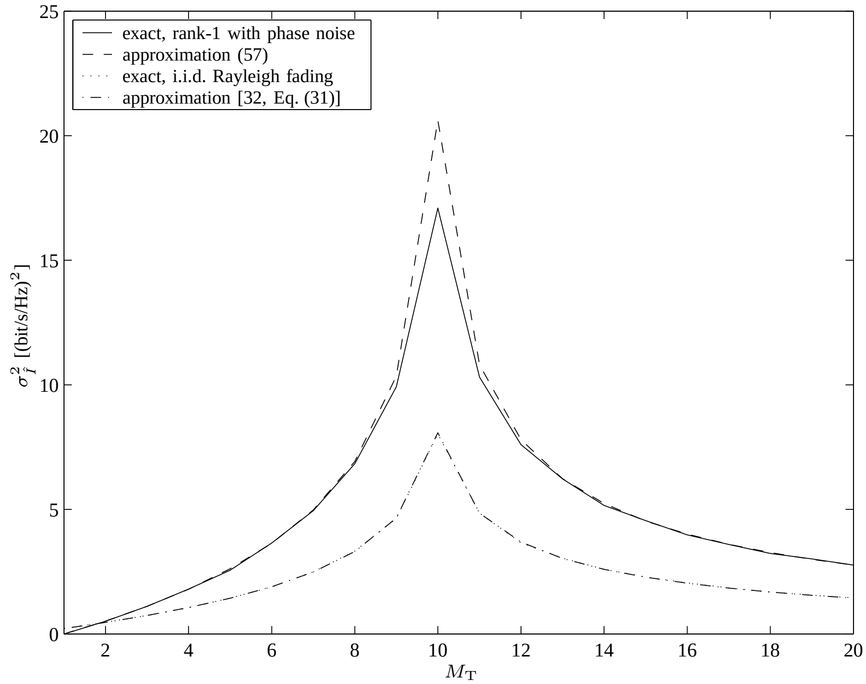

Fig. 5. Exact variance of mutual information and analytic approximation (56) at high SNR for $M_{\mathrm{R}} = 10$ as a function of M_{T} for a rank-1 physical channel with unit-modulus entries subject to 3.5° rms fully uncorrelated phase noise, and for an i.i.d. Rayleigh fading channel. Exact results are obtained through Monte Carlo simulation.

show that

$$2\sum_{i=1}^{M_{\mathrm{R}}-1} \Psi'(M_{\mathrm{T}}-i) \geq \sum_{i=0}^{M_{\mathrm{R}}-1} \Psi'(M_{\mathrm{T}}-i)\,. \tag{58}$$

Subtracting the common terms on both sides, it follows that (58) is equivalent to

$$\sum_{i=1}^{M_{\mathrm{R}}-1} \Psi'(M_{\mathrm{T}}-i) \geq \Psi'(M_{\mathrm{T}})\,.$$

The final result follows from the monotonicity property $\Psi'(M_{\mathrm{T}}-i) \geq \Psi'(M_{\mathrm{T}})$, $i \geq 1$. We note that (58) suggests that the variance of MI in the phase noise case is essentially twice that obtained for an i.i.d. Gaussian channel with the same number of transmit and receive antennas. The underlying reason lies in the fact that the individual chi-square terms in the phase noise case (cf. (44) and (45)) have half the number of degrees of freedom when compared to the Gaussian channel case. The following numerical example corroborates the factor-2 statement motivated by the inequality (58).

Numerical result: For a deterministic rank-1 physical channel $\mathbf{H} = \mathbf{g}\mathbf{h}^T$ with $|[\mathbf{g}]_i| = 1$, $\forall i$, and $|[\mathbf{h}]_i| = 1$, $\forall i$, subject to 3.5° rms fully uncorrelated phase noise, Fig. 5 shows $\sigma^2_{\hat{I}}$ according to the approximation (56) along with the exact result (obtained from Monte Carlo simulation) for $M_{\mathrm{R}} = 10$ as a function of M_{T}. We can see that the approximation is very tight for $M_{\mathrm{T}} \neq M_{\mathrm{R}}$ and predicts the location of the maximum of $\sigma^2_{\hat{I}}$ accurately. For comparison, we show σ^2_I according to the approximation [32, Eq. (31)] along with the exact result (obtained from Monte Carlo simulation) for an i.i.d. Rayleigh fading channel with the same number of transmit and receive antennas. It is clearly seen that both types of channels exhibit similar MI variance behavior as a function of M_{T}, and that $\sigma^2_{\hat{I}} \approx 2\sigma^2_I$.

VI. Conclusions

We showed that phase errors (caused by phase noise and carrier frequency offset) in time-division multiplexed switching (TDMS)-based multiple-input multiple-output (MIMO) radio channel sounders can lead to severe mutual information (MI) and capacity estimation errors. The impact of phase errors is most pronounced for low-rank physical channels where overestimation by several hundred percent can occur. A detailed analysis of the rank-1 physical MIMO channel revealed that realistic phase noise properties lead to a decorrelation of the channel matrix and result in a full-rank effective channel. Using matrix differential calculus and matrix-variate Wirtinger calculus, we characterized the sensitivity of MI w.r.t. phase errors.

In the light of the main findings of this paper, the results obtained through MIMO channel measurement campaigns using TDMS-based MIMO channel sounders should be interpreted with great care. To the best of our knowledge, the large majority of commercially available MIMO channel sounders is TDMS-based. In particular, measurement results reporting the absence of pin-hole or key-hole channels [3], [4], seem questionable unless a channel sounder with separate radio frequency (RF) chains for the individual transmit and/or receive antenna elements is used.

A discussion on hardware aspects of TDMS-based MIMO channel sounders along with extensive measurement results can be found in [1], where we also derive the abstract signal and system model used in this paper and verify the underlying assumptions.

Some more insight into state of the art clock accuracies and corresponding error in mutual information can be found in [34], as well as a more in-depth analysis of ways for mitigating these errors.

Appendix A

Theorem 10 (Fang, Kotz, and Ng [10, Th. 1.4]): Let $\mathbf{x} \stackrel{\mathrm{d}}{=} D_n(\mathbf{a})$ be partitioned into k subvectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}$ and $\mathbf{a}$ into the corresponding subvectors $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(k)}$. Let y_i and b_i be, respectively, the sums of the components of $\mathbf{x}^{(i)}$ and $\mathbf{a}^{(i)}$, and set $\mathbf{z}^{(i)} = \mathbf{x}^{(i)}/y_i$. The following statements hold:

i) The vectors $\mathbf{z}^{(i)}, \mathbf{z}^{(i+1)}, \ldots, \mathbf{z}^{(k)}$, and $\mathbf{y} = [\, y_1 \ \ y_2 \ \ \cdots \ \ y_k \,]^T$ are statistically independent.
ii) $\mathbf{y}$ is distributed as $D_k(b_1, b_2, \ldots, b_k)$.
iii) $\mathbf{z}^{(i)}$ is distributed as $D_{n_i}(\mathbf{a}^{(i)})$, $i = 1, 2, \ldots, k$, where n_i is the number of elements in the subvectors $\mathbf{a}^{(i)}$ and $\mathbf{z}^{(i)}$.

From Theorem 10 it follows immediately that the vectors $\mathbf{x}^{(i)}$ are Dirichlet (or beta) distributed. This result can be stated formally as follows.

Theorem 11 (Fang, Kotz, and Ng [10, Th. 1.5]): If $[\, x_1 \ \ x_2 \ \ \cdots \ \ x_{n-1} \,] \stackrel{\mathrm{d}}{=} D_{n-1}(a_1, a_2, \ldots, a_{n-1}; a_n)$, then for any $k < n-1$, we have $[\, x_1 \ \ x_2 \ \ \cdots \ \ x_k \,] \stackrel{\mathrm{d}}{=} D_k(a_1, a_2, \ldots, a_k; b)$, where $b = a_{k+1} + a_{k+2} + \cdots + a_n$. In particular, for any i $(i = 1, 2, \ldots, n-1)$, we have $x_i \stackrel{\mathrm{d}}{=} \beta(a_i, a - a_i)$ with $a = \sum_{i=1}^{n} a_i$.

Lemma 3: The first derivative $\Psi'(z) = \mathrm{d}\Psi(z)/\mathrm{d}z$ of the digamma function $\Psi(z)$ defined in (2) satisfies

$$\Psi'(z) \leq \frac{1}{2}\, \Psi'\Big(\frac{z}{2}\Big).$$

Proof: Using (4), we have

$$\begin{aligned}
\Psi'\Big(\frac{x}{2}\Big) &= \sum_{p=0}^{\infty} \frac{1}{\big(p + \frac{x}{2}\big)^2} = \sum_{p=0}^{\infty} \frac{4}{(2p+x)^2} \\
&= 4\left(\frac{1}{x^2} + \frac{1}{(2+x)^2} + \frac{1}{(4+x)^2} + \cdots\right) \\
&= 4\left(\sum_{p=0}^{\infty} \frac{1}{(p+x)^2} - \sum_{p=0}^{\infty} \frac{1}{\big((2p+1)+x\big)^2}\right) \\
&= 4\,\Psi'(x) - \Psi'\Big(\frac{x+1}{2}\Big)
\end{aligned}$$

which upon noting that

$$\Psi'\Big(\frac{x}{2}\Big) \geq \Psi'\Big(\frac{x+1}{2}\Big)$$

completes the proof. ■

Lemma 4: The eigenvalues of the matrix

$$\mathrm{Re}(\mathbf{A}_i) = \mathbf{I}_{M_\mathrm{T}} - \sum_{n=1}^{i-1} \frac{\mathrm{Re}(\mathbf{s}_n^{\perp} \mathbf{s}_n^{\perp H})}{\|\mathbf{s}_n^{\perp}\|^2}, \qquad i = 2, 3, \ldots, M_\mathrm{R}$$

with $\mathbf{A}_i$ and $\mathbf{s}_n^{\perp}$ defined in (33) and $\mathbf{s}_n$ defined through $\mathbf{S} = [\, \mathbf{s}_1 \ \ \mathbf{s}_2 \ \ \ldots \ \ \mathbf{s}_{M_\mathrm{R}} \,]^T = -j\boldsymbol{\Sigma} + \widetilde{\boldsymbol{\Phi}}$ with $\boldsymbol{\Sigma}$ as in (31), are given by $(k = 1, 2, \ldots, M_\mathrm{T})$

$$\{\sigma_k^{(i)}\} = \{\underbrace{1, \ldots, 1}_{M_\mathrm{T}-i}, \underbrace{0, \ldots, 0}_{i-2}, \eta^{(i)}, 1 - \eta^{(i)}\}$$

where $\eta^{(i)} = \eta^{(i)}(\mathbf{s}_1^{\perp}, \mathbf{s}_2^{\perp}, \ldots, \mathbf{s}_{i-1}^{\perp}) \in [0, 1]$.

Proof: We start by writing

$$\begin{aligned}
\mathrm{Re}(\mathbf{A}_i) &= \mathbf{I}_{M_\mathrm{T}} - \sum_{n=1}^{i-1} \frac{\mathrm{Re}(\mathbf{s}_n^{\perp})\,\big(\mathrm{Re}(\mathbf{s}_n^{\perp})\big)^T}{\|\mathbf{s}_n^{\perp}\|^2} \\
&\quad - \sum_{n=1}^{i-1} \frac{\mathrm{Im}(\mathbf{s}_n^{\perp})\,\big(\mathrm{Im}(\mathbf{s}_n^{\perp})\big)^T}{\|\mathbf{s}_n^{\perp}\|^2} \\
&= \mathbf{I}_{M_\mathrm{T}} - \mathbf{G}_i \mathbf{G}_i^T
\end{aligned}$$

where

$$\mathbf{G}_i = \big[\, \mathrm{Re}(\mathbf{N}_i) \ \ \mathrm{Im}(\mathbf{N}_i) \,\big] \qquad \text{with}$$
$$\mathbf{N}_i = \left[\, \frac{\mathbf{s}_1^{\perp}}{\|\mathbf{s}_1^{\perp}\|} \ \ \frac{\mathbf{s}_2^{\perp}}{\|\mathbf{s}_2^{\perp}\|} \ \ \cdots \ \ \frac{\mathbf{s}_{i-1}^{\perp}}{\|\mathbf{s}_{i-1}^{\perp}\|} \,\right].$$

By slight abuse of notation, in the remainder of this proof, we let $\lambda_k(\mathbf{X})$ denote the unordered eigenvalues of the matrix $\mathbf{X}$. It follows that $\sigma_k^{(i)} = 1 - \lambda_k(\mathbf{G}_i \mathbf{G}_i^T)$, $k = 1, 2, \ldots, M_\mathrm{T}$. Invoking Lemma 5, we can conclude that the $2(i-1)$ eigenvalues $\lambda_k(\mathbf{G}_i^T \mathbf{G}_i)$ are given by

$$\begin{aligned}
&\bigg\{\frac{1}{2} + \frac{1}{2}\sqrt{\mu_1^{(i)}}, \frac{1}{2} - \frac{1}{2}\sqrt{\mu_1^{(i)}}, \ldots, \\
&\quad \frac{1}{2} + \frac{1}{2}\sqrt{\mu_{i-1}^{(i)}}, \frac{1}{2} - \frac{1}{2}\sqrt{\mu_{i-1}^{(i)}}\bigg\}
\end{aligned} \tag{59}$$

with $\mu_k^{(i)} \in \mathbb{R}$ $(k = 1, 2, \ldots, i-1)$. Therefore, when paired properly, the $\lambda_k(\mathbf{G}_i^T \mathbf{G}_i)$ pairwise add up to 1. This property will next allow us to show that $\mathbf{G}_i^T \mathbf{G}_i$ has $i-2$ eigenvalues equal to 1, $i-2$ eigenvalues equal to 0, and one pair of eigenvalues given by $\{\eta^{(i)}, 1 - \eta^{(i)}\}$. We start by noting that, using (33), it follows for $i = 2, 3, \ldots, M_\mathrm{R}$ that (recall that $\mathbf{s}_i \in \mathbb{R}^{M_\mathrm{T}}$)

$$\begin{aligned}
&\mathrm{Im}\bigg(\frac{\mathbf{s}_i^{\perp}}{\|\mathbf{s}_i^{\perp}\|}\bigg) \\
&\quad = -\sum_{n=1}^{i-1} \frac{\mathrm{Im}(\mathbf{s}_n^{\perp} \mathbf{s}_n^{\perp H})}{\|\mathbf{s}_n^{\perp}\|^2}\, \frac{\mathbf{s}_i}{\|\mathbf{s}_i^{\perp}\|} \\
&\quad = -\sum_{n=1}^{i-1} \frac{\mathrm{Im}(\mathbf{s}_n^{\perp})\,\big(\mathrm{Re}(\mathbf{s}_n^{\perp})\big)^T - \mathrm{Re}(\mathbf{s}_n^{\perp})\,\big(\mathrm{Im}(\mathbf{s}_n^{\perp})\big)^T}{\|\mathbf{s}_n^{\perp}\|^2}\, \frac{\mathbf{s}_i}{\|\mathbf{s}_i^{\perp}\|} \\
&\quad = \sum_{n=1}^{i-1} \big(\mathrm{Im}(\mathbf{s}_n^{\perp})\, \xi_n + \mathrm{Re}(\mathbf{s}_n^{\perp})\, \zeta_n\big)
\end{aligned} \tag{60}$$

$$\begin{aligned}
&\mathrm{Re}\bigg(\frac{\mathbf{s}_i^{\perp}}{\|\mathbf{s}_i^{\perp}\|}\bigg) \\
&\quad = \bigg(\mathbf{I}_{M_\mathrm{T}} - \sum_{n=1}^{i-1} \frac{\mathrm{Re}(\mathbf{s}_n^{\perp})\,\big(\mathrm{Re}(\mathbf{s}_n^{\perp})\big)^T}{\|\mathbf{s}_n^{\perp}\|^2} \\
&\qquad - \sum_{n=1}^{i-1} \frac{\mathrm{Im}(\mathbf{s}_n^{\perp})\,\big(\mathrm{Im}(\mathbf{s}_n^{\perp})\big)^T}{\|\mathbf{s}_n^{\perp}\|^2}\bigg) \frac{\mathbf{s}_i}{\|\mathbf{s}_i^{\perp}\|}
\end{aligned}$$

$$= \frac{\mathbf{s}_i}{\|\mathbf{s}_i^\perp\|} + \sum_{n=1}^{i-1} \big(\mathrm{Re}(\mathbf{s}_n^\perp)\,\xi_n + \mathrm{Im}(\mathbf{s}_n^\perp)\,\zeta_n\big) \tag{61}$$

with $\xi_n, \zeta_n \in \mathbb{R}$. The significance of (60) and (61) is that it can be used to show that $\mathrm{r}(\mathbf{G}_i) = i$. More specifically, starting with $i = 2$, we note that $\mathbf{G}_2 = [\,\mathrm{Re}(\mathbf{s}_1/\|\mathbf{s}_1\|) \;\; \mathrm{Im}(\mathbf{s}_1/\|\mathbf{s}_1\|)\,]$ has rank 2 w.p. 1 as will be shown first. Noting that $\mathrm{Re}(\mathbf{s}_1) \overset{\mathrm{d}}{=} \mathcal{N}(\mathbf{0}, \sigma_\varphi^2\mathbf{I})$ and $\mathrm{Im}(\mathbf{s}_1) = [\,-M_\mathrm{T}M_\mathrm{R} \;\; \mathbf{0}_{1,M_\mathrm{T}-1}\,]^T$, it follows that $\|\mathbf{s}_1\| > 0$ w.p. 1 and hence $r(\mathbf{G}_2) = r([\,\mathrm{Re}(\mathbf{s}_1) \;\; \mathrm{Im}(\mathbf{s}_1)\,])$. By definition, $\mathrm{Re}(\mathbf{s}_1)$ and $\mathrm{Im}(\mathbf{s}_1)$ are linearly independent w.p. 1 and hence $r(\mathbf{G}_2) = 2$ w.p. 1. Now with each increase in i, two vectors are added to $\mathbf{G}_i$, where one, namely $\mathrm{Im}(\mathbf{s}_{i-1}^\perp/\|\mathbf{s}_{i-1}^\perp\|)$ by (60), is a linear combination of the vectors already in $\mathbf{G}_i$, and the other one, namely $\mathrm{Re}(\mathbf{s}_{i-1}^\perp/\|\mathbf{s}_{i-1}^\perp\|)$ by (61), is a linear combination of the vectors already in $\mathbf{G}_i$ and the vector $\mathbf{s}_{i-1}/\|\mathbf{s}_{i-1}^\perp\|$. We can therefore conclude that the $2(i-1)\times 2(i-1)$ matrix $\mathbf{G}_i^T\mathbf{G}_i$ has at most i nonzero eigenvalues. The remaining $i-2$ eigenvalues are equal to zero. By (59) we must therefore have $i-2$ eigenvalues which are equal to 1. Since we have $2i-2$ eigenvalues in total, it follows that there is one pair of eigenvalues of the form $\{\eta^{(i)}, 1-\eta^{(i)}\}$. Finally, noting that $\eta^{(i)}$ and $1-\eta^{(i)}$ must be real-valued and positive, we can conclude that $\eta^{(i)} \in [0,1]$. ■

Lemma 5: Given an orthonormal set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ with $n \le N$, the eigenvalues of the $2n \times 2n$ matrix $\mathbf{Y}^T\mathbf{Y}$ with

$$\mathbf{Y} = \big[\,\mathrm{Re}(\mathbf{X}) \;\; \mathrm{Im}(\mathbf{X})\,\big]$$

where

$$\mathbf{X} = [\,\mathbf{x}_1 \;\; \mathbf{x}_2 \;\; \cdots \;\; \mathbf{x}_n\,]$$

are given by

$$\begin{aligned}\Bigg\{&\frac{1+\sqrt{\lambda_1(\mathbf{A})}}{2}, \frac{1-\sqrt{\lambda_1(\mathbf{A})}}{2},\\ &\frac{1+\sqrt{\lambda_2(\mathbf{A})}}{2}, \frac{1-\sqrt{\lambda_2(\mathbf{A})}}{2}, \ldots,\\ &\frac{1+\sqrt{\lambda_n(\mathbf{A})}}{2}, \frac{1-\sqrt{\lambda_n(\mathbf{A})}}{2}\Bigg\}\end{aligned} \tag{62}$$

where $\mathbf{A} = \mathbf{X}^T\mathbf{X}\mathbf{X}^H\mathbf{X}^*$.

Proof: We start by noting that

$$\mathbf{Y}' = \sqrt{\frac{1}{2}}\,[\,\mathbf{X} \;\; \mathbf{X}^*\,]$$

satisfies

$$\mathbf{Y}' = \mathbf{Y}\underbrace{\left(\sqrt{\frac{1}{2}}\begin{bmatrix} 1 & 1 \\ j & -j\end{bmatrix} \otimes \mathbf{I}_n\right)}_{\mathbf{U}}$$

where $\mathbf{U}$ is unitary. We can therefore conclude that

$$\begin{aligned}\lambda_i(\mathbf{Y}'^H\mathbf{Y}') &= \frac{1}{2}\,\lambda_i\left(\begin{bmatrix}\mathbf{X}^H \\ \mathbf{X}^T\end{bmatrix}[\,\mathbf{X} \;\; \mathbf{X}^*\,]\right)\\ &= \lambda_i(\mathbf{U}^H\mathbf{Y}^H\mathbf{Y}\mathbf{U}) = \lambda_i(\mathbf{Y}^H\mathbf{Y}).\end{aligned}$$

Next, we note that $\mathbf{X}^H\mathbf{X} = \mathbf{X}^T\mathbf{X}^* = \mathbf{I}_n$ due to the orthonormality of the $\mathbf{x}_n$, and hence the eigenvalues of $\mathbf{Y}^H\mathbf{Y}$ are given by the solutions of the characteristic equation

$$\mathrm{p}(\mu) = \det\left(\begin{bmatrix}(1-2\mu)\mathbf{I} & \mathbf{X}^H\mathbf{X}^* \\ \mathbf{X}^T\mathbf{X} & (1-2\mu)\mathbf{I}\end{bmatrix}\right).$$

For a Hermitian matrix $\mathbf{S}$ partitioned according to

$$\mathbf{S} = \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{C}\end{bmatrix}$$

where $\mathbf{A}$ and $\mathbf{C}$ are square matrices, we have from the Schur complement formula [18, Sec. 0.8.5]

$$\det(\mathbf{S}) = \det(\mathbf{A})\det(\mathbf{C} - \mathbf{B}^*\mathbf{A}^{-1}\mathbf{B}).$$

It therefore follows that

$$p(\mu) = \det\big((1-2\mu)^2\mathbf{I} - \mathbf{X}^T\mathbf{X}\mathbf{X}^H\mathbf{X}^*\big). \tag{63}$$

The solution of (63) yields the eigenvalues of $\mathbf{Y}^H\mathbf{Y}$ as (62), which completes the proof. ■

Note that if $n \ge N/2$, $\mathbf{Y}$ must have $2n-N$ eigenvalues equal to 0 and therefore by (62) also $2n-N$ eigenvalues equal to 1 so that (62) can be refined as

$$\begin{aligned}\Bigg\{&\frac{1+\sqrt{\lambda_1(\mathbf{A})}}{2}, \frac{1-\sqrt{\lambda_1(\mathbf{A})}}{2},\\ &\frac{1+\sqrt{\lambda_2(\mathbf{A})}}{2}, \frac{1-\sqrt{\lambda_2(\mathbf{A})}}{2}, \ldots,\\ &\frac{1+\sqrt{\lambda_{N-n}(\mathbf{A})}}{2}, \frac{1-\sqrt{\lambda_{N-n}(\mathbf{A})}}{2},\\ &\underbrace{1,\ldots,1}_{2n-N}, \underbrace{0,\ldots,0}_{2n-N}\Bigg\}.\end{aligned}$$

Theorem 12: Given the conditional (on η) RV[9] $Z(\eta) \overset{\mathrm{d}}{=} \chi^2_{1,\sigma_\varphi^2\eta} + \chi^2_{1,\sigma_\varphi^2(1-\eta)}$, where the two chi-square distributed terms are independent, and a constant $r \in \mathbb{R}^+$, we have

$$\mathbb{E}\big\{\log\big(r + Z(0)\big)\big\} \le \mathbb{E}\big\{\log\big(r + Z(\eta)\big)\big\}, \quad 0 \le \eta \le 1. \tag{64}$$

Proof: We start by noting that $Z(\eta) \overset{\mathrm{d}}{=} Z(1-\eta)$ for $0 \le \eta \le 1$ so that it suffices to prove (64) for $0 \le \eta \le 1/2$. We shall also need the properties $Z(0) \overset{\mathrm{d}}{=} Z(1) \overset{\mathrm{d}}{=} \chi^2_{1,\sigma_\varphi^2}$ and $Z(1/2) \overset{\mathrm{d}}{=} \chi^2_{2,\sigma_\varphi^2/2}$. Noting that

$$\begin{aligned}\log\big(r + Z(\eta)\big) &= \log(r) + \log\big(1 + Z(\eta)/r\big)\\ &= \log(r) + \log\big(1 + \tilde{Z}(\eta)\big)\end{aligned}$$

where $\tilde{Z}(\eta) \overset{\mathrm{d}}{=} \chi^2_{1,\tilde{\sigma}_\varphi^2\eta} + \chi^2_{1,\tilde{\sigma}_\varphi^2(1-\eta)}$ and $\tilde{\sigma}_\varphi^2 = \sigma_\varphi^2/r$, it follows that proving (64) for $r = 1$ is sufficient.

The cdf of $Y(\xi) = \chi^2_{1,\xi}$ is [11, Eq. 26.4.19, Eq. 6.5.16, Eq. 6.1.8]

$$\mathrm{F}_{Y(\xi)}(y) = \mathrm{erf}\left(\sqrt{\frac{y}{2\xi}}\right) \tag{65}$$

where $\mathrm{erf}(x) = (2/\sqrt{\pi})\int_0^x \exp(-t^2)\,\mathrm{d}t$ denotes the error

[9]For the sake of simplicity of notation, we committed an abuse of notation here, as in the main body (cf. Theorem 5) we used the symbol $Z(\eta)$ to denote the unconditional RV $Z(\eta) = \sigma_\varphi^2(\eta X_1 + (1-\eta)X_2)$.

function. The inverse function corresponding to (65) is given by $\mathrm{F}_{Y(\xi)}^{-1}(x) = 2\xi \mathrm{erfi}^2(x)$, where $\mathrm{erfi}(x)$ denotes the inverse error function. Using the inverse method of generating random deviates [11, Sec. 26.8], i.e., for a uniformly distributed RV $U \in [0,1]$ we have $\chi^2_{1,\xi} \overset{\mathrm{d}}{=} \mathrm{F}_{Y(\xi)}^{-1}(U)$, together with the independence of the two terms in $Z(\eta) \overset{\mathrm{d}}{=} \chi^2_{1,\sigma_\varphi^2\eta} + \chi^2_{1,\sigma_\varphi^2(1-\eta)}$, we can now express $Z(\eta)$ in terms of two independent uniformly distributed RVs $U_1 \in [0,1]$ and $U_2 \in [0,1]$ as

$$Z(\eta) \overset{\mathrm{d}}{=} 2\sigma^2\eta\,\mathrm{erfi}^2(U_1) + 2\sigma^2(1-\eta)\,\mathrm{erfi}^2(U_2). \tag{66}$$

To prove the Theorem, it suffices to show that

$$\frac{\partial}{\partial\eta}\mathbb{E}\big\{\log\big(1+Z(\eta)\big)\big\} \geq 0, \quad 0 \leq \eta \leq 1/2. \tag{67}$$

Inserting (66) in (67), we get

$$\begin{aligned}
&\frac{\partial}{\partial\eta}\mathbb{E}\big\{\log\big(1+Z(\eta)\big)\big\} \\
&\quad= \frac{\partial}{\partial\eta}\int_0^1\int_0^1 \log\big(1+2\sigma^2\eta\,\mathrm{erfi}^2(u_1) \\
&\qquad\qquad + 2\sigma^2(1-\eta)\,\mathrm{erfi}^2(u_2)\big)\,\mathrm{d}u_1\mathrm{d}u_2 \\
&\quad\overset{\text{(a)}}{=} \log(e) \\
&\qquad\cdot \iint_0^1 \frac{2\sigma^2\big(\mathrm{erfi}^2(u_1)-\mathrm{erfi}^2(u_2)\big)}{1+2\sigma^2(\eta\,\mathrm{erfi}^2(u_1)+(1-\eta)\,\mathrm{erfi}^2(u_2))}\,\mathrm{d}u_1\mathrm{d}u_2 \\
&\quad= \log(e)\int_0^1\int_0^1 f_1(u_1,u_2)f_2(u_1,u_2)\,\mathrm{d}u_1\mathrm{d}u_2
\end{aligned} \tag{68}$$

where in (a) we interchanged expectation and differentiation (noting that the integrand is continuous for all $\eta \in [0,1]$), and we set

$$\begin{aligned}
f_1(u_1,u_2) &= \mathrm{erfi}^2(u_1) - \mathrm{erfi}^2(u_2) \\
f_2(u_1,u_2) &= \big(1/(2\sigma^2) + \eta\,\mathrm{erfi}^2(u_1) + (1-\eta)\,\mathrm{erfi}^2(u_2)\big)^{-1}.
\end{aligned}$$

Note that $f_1(u_1,u_2)$ is negative symmetric, i.e., $f_1(u_1,u_2) = -f_1(u_2,u_1)$. Furthermore, the monotonicity of $\mathrm{erfi}(u)$ implies that for $u_1 \geq u_2$ we have $f_1(u_1,u_2) \geq 0$, and for $u_1 \leq u_2$ it holds that $f_1(u_1,u_2) \leq 0$. We will now exploit these properties of $f_1(u_1,u_2)$ to complete the proof and start by rewriting the integral in (68) according to

$$\begin{aligned}
&\int_{u_1=0}^1\int_{u_2=0}^1 f_1(u_1,u_2)f_2(u_1,u_2)\,\mathrm{d}u_1\,\mathrm{d}u_2 \\
&\quad= \int_{v=0}^1\bigg(\int_{u_1=v}^1 f_1(u_1,v)f_2(u_1,v)\,\mathrm{d}u_1 \\
&\qquad\qquad + \int_{u_2=v}^1 f_1(v,u_2)f_2(v,u_2)\,\mathrm{d}u_2\bigg)\mathrm{d}v \\
&\quad= \int_{v=0}^1\int_{u=v}^1 f_1(u,v)\big(f_2(u,v)-f_2(v,u)\big)\,\mathrm{d}u\,\mathrm{d}v
\end{aligned} \tag{69}$$

where $f_1(u,v) \geq 0$ in the entire range of integration since $u \geq v$. Finally, we will show that $f_2(u,v) - f_2(v,u) \geq 0$, which by (69) then implies (67). Straightforward manipulations reveal that the condition $f_2(u,v) - f_2(v,u) \geq 0$ is equivalent to

$$\eta\,\mathrm{erfi}^2(u) + (1-\eta)\,\mathrm{erfi}^2(v) \leq \eta\,\mathrm{erfi}^2(v) + (1-\eta)\,\mathrm{erfi}^2(u)$$

$$(2\eta-1)\,\mathrm{erfi}^2(u) \leq (2\eta-1)\,\mathrm{erfi}^2(v)$$

and hence by $0 \leq \eta \leq 1/2$ to

$$\mathrm{erfi}^2(u) \geq \mathrm{erfi}^2(v)$$

which is satisfied for $u \geq v$ (i.e., over the entire range of integration) because $\mathrm{erfi}(x)$ is nondecreasing. ■

Acknowledgment

We would like to thank Vinko Erceg and Pieter van Rooyen for helping both plan and making possible the measurement campaign that led to the discovery of the phase noise problem in TDMS-based MIMO channel sounding.

References

[1] D. S. Baum and H. Bölcskei, "The effect of phase errors in MIMO channel sounding," *IEEE Trans. Antennas Propag.*, 2010, to be submitted.

[2] G. Matz, A. F. Molisch, F. Hlawatsch, M. Steinbauer, and I. Gaspard, "On the systematic measurement error of correlative mobile radio channel sounders," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 808–821, May 2002.

[3] D. Gesbert, H. Bölcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[4] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 361–368, Apr. 2002.

[5] M. J. Gans, N. Amitay, Y. S. Yeh, H. Xu, T. C. Damen, R. A. Valenzuela, T. Sizer, R. Storz, D. Taylor, W. M. MacDonald, C. Tran, and A. Adamiecki, "Outdoor BLAST measurement system at 2.44 GHz: Calibration and initial results," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 570–583, Apr. 2002.

[6] P. Kyritsi, R. A. Valenzuela, and D. C. Cox, "Channel and capacity estimation errors," *IEEE Commun. Lett.*, vol. 6, no. 12, pp. 517–519, Dec. 2002.

[7] S. Loyka and A. Kouki, "On MIMO channel capacity, correlations, and keyholes: Analysis of degenerate channels," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1886–1888, Dec. 2002.

[8] P. Almers, F. Tufvesson, and A. F. Molisch, "Measurement of keyhole effect in a wireless multiple-input multiple-output (MIMO) channel," *IEEE Commun. Lett.*, vol. 7, no. 8, pp. 373–375, Aug. 2003.

[9] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd ed., ser. Wiley Series in Probability & Statistics. Chichester, UK: Wiley-VCH, 1999.

[10] K.-T. Fang, S. Kotz, and K.-W. Ng, *Symmetric Multivariate and Related Distributions*, ser. Monographs on Statistics and Applied Probability. London, UK and New York, NY: Chapman and Hall, 1990.

[11] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York, NY: Dover, 1972.

[12] G. Sommerkorn, A. Richter, R. S. Thomä, and W. Wirnitzer, "Antenna multiplexing and time alignment for MIMO channel sounding," in *Proc. Int. Union of Radio Science (URSI) General Assembly (GA)*, Maastricht, The Netherlands, Aug. 2002, CP.P.28, paper no. 1879.

[13] A. Hajimiri and T. H. Lee, *The Design of Low Noise Oscillators*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1999.

[14] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[15] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[16] R. A. Wooding, "The multivariate distribution of complex normal variables," *Biometrika*, vol. 43, no. 1/2, pp. 212–215, Jun. 1956.

[17] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," Lecture Supplement to ECE275A, Ver. ECE275CG-F05v1.3c, University of California, San Diego, USA, 2005. [Online]. Available: http://dsp.ucsd.edu/~kreutz/PEI05.html

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[19] M. L. Eaton and M. D. Perlman, "The non-singularity of generalized sample covariance matrices," *Ann. Stat.*, vol. 1, no. 4, pp. 710–717, Jul. 1973.

[20] M. Capiński and P. E. Kopp, *Measure, Integral and Probability*, 2nd ed. London, UK: Springer, 1999.

[21] T. McKelvey, H. Akçay, and L. Ljung, "Subspace-based multivariable system identification from frequency response data," *IEEE Trans. Autom. Control*, vol. 41, no. 7, pp. 960–979, Jul. 1996.

[22] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. John Wiley & Sons, 2003.

[23] A. M. Mathai, *An Introduction to Geometrical Probability: Distributional Aspects with Applications*. Newark, NJ: Gordon and Breach, 1999.

[24] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, MIT, Cambridge, MA, 1989.

[25] A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables: Theory and Applications*. New York, NY: M. Dekker, 1992.

[26] D. S. Baum and H. Bölcskei, "Impact of phase noise on MIMO channel measurement accuracy," in *Proc. IEEE Veh. Technol. Conf. (VTC) Fall*, vol. 3, Los Angeles, CA, Sep. 2004, pp. 1614–1618.

[27] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*. Kluwer Academic Publishers, 2002.

[28] J. C. Hansen and H. Bölcskei, "A geometrical investigation of the rank-1 Ricean MIMO channel at high SNR," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Chicago, IL, Jun. 2004, p. 64.

[29] G. W. Stewart, "The efficient generation of random orthogonal matrices with an application to condition estimators," *SIAM J. Numerical Anal.*, vol. 17, no. 3, pp. 403–409, Jun. 1980.

[30] Q. T. Zhang, X. W. Cui, and X. M. Li, "Very tight capacity bounds for MIMO-correlated Rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 681– 688, Mar. 2005.

[31] P. M. Lee, *Bayesian Statistics: An Introduction*, 2nd ed. London, UK: Arnold, New York, NY: Wiley, 1997.

[32] Ö. Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.

[33] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[34] D. S. Baum, "Information-theoretic analysis of a class of MIMO channel measurement devices," Ph.D. dissertation, ETH Zurich, Zürich, Switzerland, 2007, No. 17395. [Online]. Available: http://e-collection.ethbib.ethz.ch/view/eth:30410