

Massive But Few Active MIMO

Dushyantha A. Basnayaka, *Member, IEEE*, Marco Di Renzo, *Member, IEEE*, and Harald Haas, *Member, IEEE*

Abstract—In this paper, an emerging wireless communication concept, which is termed as spatial modulation (SM) for large-scale multiple-input-multiple-output (MIMO), is considered. The results show that from the information-theoretic perspective, SM achieves capacity comparable to the open-loop MIMO capacity, although a subset of transmit antennas is activated in every channel use because both the channel coefficients and the input symbols carry information in SM. As a result, SM compensates the loss of information capacity due to a subset of antennas being activated by modulating information in the antenna index; therefore, the total information rate remains high. In particular, an upper bound for the capacity of SM is derived in closed form, and it is shown analytically that this upper bound is almost certainly achievable in the massive MIMO regime. Moreover, it is shown that the upper bound is achievable with no channel state information at the transmitter (CSIT) but with channel distribution information (CDI) at the transmitter (CDIT). The optimum transmission strategy should adapt the channel input distribution to fading using CDI such as the K factor in Rician fading or the shape parameter m in Nakagami- m fading.

Index Terms—Capacity, channel distribution information (CDI), channel state information (CSI), channel state information at the transmitter (CSIT), massive multiple-input multiple-output (MIMO), spatial modulation (SM).

I. INTRODUCTION

MULTIPLE-INPUT MULTIPLE-OUTPUT (MIMO) is a technology for increasing the link capacity and/or reliability in modern communication systems. Multielement antennas are used in many scenarios such as point-to-point links [1], [2], multiuser links [3], and macrodiversity links [4]. MIMO can achieve higher throughput because multiple independent spatial data streams can be transmitted in the same time and frequency resource, and one of the key enablers for MIMO operation is the rich scattering environment between transmit and receive antennas [5]. The success of MIMO lies in the fact that receivers can successfully separate the multiple data streams transmitted by the transmitter with the assistance of channel state information (CSI) at the receiver. It has been shown with

massive MIMO that CSI at the transmitter (CSIT) can also be used to increase the system capacity [6]. The widely used transmission scheme for massive MIMO is multiuser MIMO with linear precoding [7]. It has been shown that a massive antenna array can be used to simultaneously serve a large number of single-antenna users by sharing multiplexing gain among them. The uncorrelated noise and small-scale fading could be eliminated due to the averaging (hardening) effect. The required transmitted energy per bit decreases as the number of base-station antennas increases [8]. However, CSI overhead, mobility, and RF power consumption can cause problems [7]. In particular, acquiring CSIT in frequency-division duplexing system is a costly process. It has several phases that include estimating CSI at the receiver and feeding back appropriate information to the transmitter [9]. The feedback cost increases with the number of constituent links in the MIMO links. Therefore, on the one hand, there are considerable interests in finding other energy-efficient transmission schemes for massive MIMO with no CSIT. On the other hand, one of the main challenges for future wireless communication systems is the increasingly wide range of applications with varying requirements and characteristics [10]. The future wireless communication systems should therefore be smart, heterogeneous, and adaptive to time-varying channel conditions [11]. In the future, a wide spectrum of physical-layer (PHY) transmission schemes are required to serve different users in different scenarios [12]. Therefore, instead of relying on one PHY-layer transmission scheme, it is essential to have different schemes with diverse characteristics. There are potential transmission schemes for MIMO links in the absence of CSIT. Among them, stand-alone space-time (ST) code and spatial multiplexing (SMX) are prominent [13], [14]. In SMX, many independent data streams as the number of transmit antennas are transmitted in a single use of the channel, and it has been shown to have a reasonable decoding complexity [13]. In ST coding, redundancy is added to the transmit symbol to provide multiple independent replicas of transmit data symbols to the receiver.

In this paper, a simple information-based antenna switching technique termed spatial modulation (SM) is considered as a possible transceiver solution for MIMO links with no CSIT [15]–[17]. As explained in Section II, SM modulates information in both the signal constellation and the antenna index. In addition to its simplicity, the ability to control the number of transmit RF chains in SM can be very important from transmit energy efficiency perspective. In contrast to SMX and ST coding, SM uses a subset of transmit antennas for data transmission, and data is also modulated in the antenna index. The operation and performance of SM are now well understood [16]. The optimum detection of SM is investigated in [18], and the average bit error probability in different fading scenarios

Manuscript received February 13, 2015; revised June 30, 2015 and August 25, 2015; accepted September 19, 2015. Date of publication October 14, 2015; date of current version September 15, 2016. The work of D. A. Basnayaka and H. Haas was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Established Career Fellowship Grant EP/K008757/1. The review of this paper was coordinated by Dr. H. Lin.

D. A. Basnayaka and H. Haas are with the Institute for Digital Communication, University of Edinburgh, Edinburgh EH9 3FG, U.K. (e-mail: d.basnayaka@ed.ac.uk; h.haas@ed.ac.uk).

M. Di Renzo is with the Laboratoire des Signaux et Systèmes, Unité Mixte de Recherche 8506, Centre National de la Recherche Scientifique-École Supérieure d'Électricité-Université Paris-Sud XI, 91192 Gif-sur-Yvette, France (e-mail: marco.direnzo@lss.supelec.fr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVT.2015.2490548

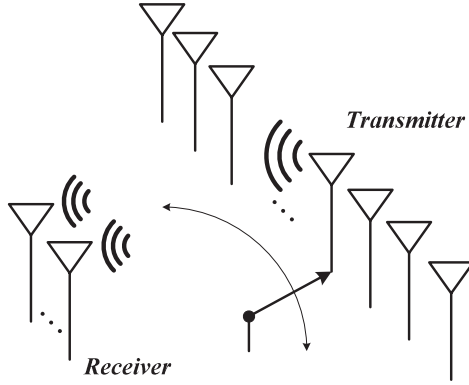


Fig. 1. Large-scale MIMO system with a fewer number of receive antennas in comparison with the number of transmit antennas.

is analyzed in [19]. The effect of channel estimation error and the effect of antenna switching on the bandwidth efficiency are investigated in [20] and [21], respectively. The channel estimation of SM is considered in [22] and [23], and energy efficiency of SM is studied in [24] and [25].

The main objective of this paper is to investigate the potential gain of SM in the massive MIMO regime with no CSIT. In particular, a MIMO scenario with a large number of transmit antennas and a relatively fewer number of receive antennas is considered, as shown in Fig. 1. This scenario typically occurs in cellular downlink communication from a base station to a user terminal. On the one hand, transmission schemes such as vertical Bell laboratories layered space–time (V-BLAST) [13] could be employed with random beamforming [26], and on the other hand, stand-alone ST code could also be used for such MIMO systems. Since both V-BLAST and ST code transmit using all transmit antennas, their energy efficiency may not scale well with the number of transmit antennas due to the large number of power amplifiers if the RF power consumption is also considered [27].

Firstly, we consider the information-theoretic capacity of SM in additive white Gaussian noise (AWGN) channels and fading. Unlike traditional MIMO links, the wireless channel itself carries information; therefore, capacity analyses tend to be more challenging. Capacity analysis has been reported in several prior works [28], [29]. The approach in many reported research is analyzing the information conveyed in the antenna index and signal constellation separately. Moreover, their capacity analysis does not have a closed-form solution, which could be used for further optimization. However, in this paper a different approach is used in which the system is modeled as two independent sources of information, constellation, and antenna index with multiplicative interaction. This model allows the calculation of the capacity in a more straightforward manner. The optimal channel input distribution for SM is analytically obtained. Then, the theoretical capacity results are compared with mutual information (MI) results for SM with suboptimal channel inputs, i.e., practical modulation schemes such as quadrature amplitude modulation (QAM). The main contributions of this paper are summarized as follows.

- An upper bound for the capacity of SM is derived in closed form, and it is shown analytically that this upper

bound is almost certainly achievable in the massive MIMO regime. Moreover, the capacity upper bound is achievable with no CSIT, but channel distribution information at the transmitter (CDIT). The optimum transmission strategy should adapt the channel input distribution to fading using channel distribution information (CDI) such as the K factor in Rician fading or the shape parameter m in Nakagami fading.

- Unlike in conventional MIMO, Gaussian input distribution is not in general optimal for SM MIMO, but in some undesirable fading conditions, it can be approximately optimal, e.g., high transmit correlation and high K factor scenarios.
- In Rayleigh fading as $N \rightarrow \infty$, 4-QAM is shown to be optimal. The practical implications of this result are that, in uncorrelated Rayleigh fading, transmitting as much information as possible in the antenna index and keeping the constellation order as low as possible are desirable.
- Antenna sets should be selected in such a way that the i th data stream is transmitted from all antennas. For example, if there are four transmit antennas, and two transmit antennas are supposed to be active in each channel use, (1, 2), (1, 3), (1, 4), (2, 3) is less favorable than (1, 2), (3, 1), (4, 3), (2, 4).
- Although there is an optimum input distribution for SM as $N \rightarrow \infty$, simulation results show that it is not always required to signal with the optimum input distribution to achieve the capacity in finite N , but real constellations such as M -QAM or, in some cases, Gaussian inputs are also sufficient.

This paper contains the full mathematical derivations, extensive numerical results, and extended discussions, and the initial results of this paper were published in [30].

The remainder of this paper is laid out as follows. In Section II, the system model is described. The existing analysis and its drawbacks are discussed in Section III, followed by the main analysis in Section IV. The analysis is extended to SM with transmit correlation in Section V. Section VI provides numerical results, and Section VII gives the conclusion.

II. SYSTEM MODEL

A single-user point-to-point SM MIMO communication system in fading with N transmit and n_R receive antennas is considered. Here, it is assumed that $N \gg n_R$. The $\mathcal{C}^{n_R \times N}$ channel matrix $\mathbf{H}$ captures the fading between the transmit and receive antenna arrays. The channel matrix $\mathbf{H}$ contains elements h_{ik} , which represents the normalized complex channel gain between the transmit antenna k and the receive antenna i , where $\mathcal{E}\{|h_{ik}|^2\} = 1$. The SM scheme considered here activates several antennas for data transmission, and an independent data stream is transmitted from each active antenna. The set of active antennas is selected based on the information and the total number of such sets generally has to be an integer power of 2. There are $\binom{N}{T}$ possible sets for a SM system with T active antennas. It is assumed that only L sets are configured as legal antenna sets for data transmission, and T constellation symbols are directly transmitted from T active

antennas. As a result, a binary block $\mathbf{b}$ with $\log_2 L + T \log_2 M$ bits is transmitted in a single channel use. The first $\log_2 L$ bits are modulated in the antenna index, whereas the next $T \log_2 M$ bits are conventionally modulated using QAM constellation. The first $\log_2 L$ bits select which antenna set is activated in the current channel use and the next $T \log_2 M$ bits select the symbols that are transmitted by active antennas. In this paper, the set of active transmit antennas is denoted as Υ_i , where $\Upsilon_i \in \{\Upsilon_1, \dots, \Upsilon_L\}$. The modulation in the antenna index inevitably incurs switching delays; this antenna switching delay in SM transmission could cause some bandwidth expansion [21]. This bandwidth expansion can be regarded as a result of hardware limitation and is not a fundamental drawback of SM MIMO from the theoretical point of view. Therefore, its effect on spectral efficiency is omitted in this theoretical study. The received baseband signal thus becomes

$$\mathbf{r} = \sqrt{p}\mathbf{V}\mathbf{x} + \mathbf{n} \quad (1)$$

where $\mathcal{C}^{n_R \times T}$ matrix is $\mathbf{V} \in \{\tilde{\mathbf{H}}_1, \dots, \tilde{\mathbf{H}}_L\}$. The matrix $\tilde{\mathbf{H}}_i$ is a submatrix of $\mathbf{H}$, which consists of channel elements from all active antennas in the set Υ_i to all receive antennas. The $\{\tilde{\mathbf{H}}_1, \dots, \tilde{\mathbf{H}}_L\}$ is denoted the alphabet of $\mathbf{V}$. The $\mathcal{C}^{T \times 1}$ vector, $\mathbf{x} = (x_1, \dots, x_T)^T$ is the constellation data signal with average transmit power constraint $\mathcal{E}\{\mathbf{x}\mathbf{x}^H\} = (E_s/T)\mathbf{I}$. The constant p is defined to capture the average link power drop between the transmit antenna array and the receive antenna array due to shadowing and path loss. Here, transmit precoding is omitted due to the absence of CSIT; thus, independently coded data are transmitted from each active antenna, as in the case of V-BLAST transmission. For example, consider an SM system with $N = 4$, $T = 2$. There are six possible sets, but only four sets are selected as (1, 2), (1, 3), (1, 4), and (2, 3). The modulator selects one of these four sets, depending on the first two bits of the segment, and selects a 2×1 vector $\mathbf{x}$ from a 4-QAM constellation, depending on the next 4 bits if 6-bit transmission is desired. Furthermore, if more bits are desired to be transmitted, higher order signal constellations have to be employed. At the receiver, the maximum-likelihood decoder in a Gaussian noise channel is employed as follows:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{V}, \mathbf{x}} \|\mathbf{r} - \sqrt{p}\mathbf{V}\mathbf{x}\| \quad (2)$$

where $\|\cdot\|$ denotes the Euclidean vector norm.

In this paper, the fading is modeled statistically. Every channel element, h_{ik} is assumed drawn from a complex random process Φ , with finite raw moments. This complex random process consists of two independent and identically distributed (i.i.d.) real random processes (i.e., $\in \mathbb{R}$). Let $\mathcal{E}\{h_i^\ell\}$, and $\mathcal{E}\{h_Q^\ell\}$ be the ℓ th raw moments of real and imaginary random processes that make up Φ , respectively. In Rayleigh fading, h_{ik} is modeled as $h_{ik} \sim \mathcal{CN}(0, 1)$, where $\mathcal{CN}(0, 1)$ denotes a zero mean circularly symmetric complex Gaussian (ZMCSG) random variable (RV) with unit variance. Throughout this paper, CSI denotes the channel matrix $\mathbf{H}$, and CDI denotes the raw moments of the fading random process Φ . CDI for some well-known fading models are summarized in Table I.

TABLE I
CDI FOR DIFFERENT FADING MODELS

Fading	$\mathcal{E}\{h_i^\ell\} = \mathcal{E}\{h_Q^\ell\}$
Rayleigh	$(\ell-1)!! \left(\frac{1}{2}\right)^{\ell/2}$
Rice	$\left(\frac{1}{2(K+1)}\right)^{\ell/2} \frac{\partial e^{t\sqrt{K}+t^2/2}}{\partial t^\ell} \Big _{t=0}$
Nakagami-m	$\frac{\Gamma(\frac{m}{2} + \frac{\ell}{2})}{m^{\ell/2} \Gamma(\frac{m}{2})}$

A. Mutual Information

The MI achievable by arbitrary transmission schemes with practical constellations with perfect knowledge of the channel at the receiver is reviewed. Numerical simulation shall enable interesting capacity comparisons with analytical results. In summary, the objective is to simulate the MI between the channel inputs $\mathbf{V}, \mathbf{x}$ and the channel output $\mathbf{r}$. From [34], it is given by

$$I(\mathbf{V}, \mathbf{x}; \mathbf{r}|\mathbf{H}) = \mathcal{E}_{\mathbf{V}, \mathbf{x}, \mathbf{r}} \left\{ \log_2 \left(\frac{\Pr(\mathbf{r}|\mathbf{V}, \mathbf{x}, \mathbf{H})}{\Pr(\mathbf{r}|\mathbf{H})} \right) \right\} \quad (3)$$

where $\Pr(\mathbf{r}|\mathbf{V}, \mathbf{x}, \mathbf{H})$ is the conditional probability of $\mathbf{r}$ when $\mathbf{V}, \mathbf{x}$, and $\mathbf{H}$ are known. For the discrete input $\mathbf{V}, \mathbf{x}$, the MI in (3) can be expressed as

$$I(\mathbf{V}, \mathbf{x}; \mathbf{r}|\mathbf{H}) = \sum_{\mathbf{V}, \mathbf{x}} \Pr(\mathbf{V}, \mathbf{x}) \int \Pr(\mathbf{r}|\mathbf{V}, \mathbf{x}, \mathbf{H}) \times \log_2 \left(\frac{\Pr(\mathbf{r}|\mathbf{V}, \mathbf{x}, \mathbf{H})}{\Pr(\mathbf{r}|\mathbf{H})} \right) d\mathbf{r}. \quad (4)$$

A Monte Carlo estimate thereof is given by

$$I(\mathbf{V}, \mathbf{x}; \mathbf{r}|\mathbf{H}) \approx \frac{1}{n} \sum_{\mathbf{V}, \mathbf{x}} \Pr(\mathbf{V}, \mathbf{x}) \times \sum_{l=1}^n \log_2 \left(\frac{\Pr(\mathbf{r}^l|\mathbf{V}, \mathbf{x}, \mathbf{H})}{\Pr(\mathbf{r}^l|\mathbf{H})} \right) \quad (5)$$

where $\mathbf{r}^l$ with $l = 1, \dots, n$ are i.i.d. random samples drawn from $\mathbf{r}$. The conditional probability $\Pr(\mathbf{r}|\mathbf{V}, \mathbf{x}, \mathbf{H})$ for SM becomes

$$\Pr(\mathbf{r}|\mathbf{V}, \mathbf{x}, \mathbf{H}) = \frac{1}{(\pi\sigma^2)^{n_R}} e^{\|\mathbf{r} - \sqrt{p}\mathbf{V}\mathbf{x}\|^2}. \quad (6)$$

The number n is chosen to be enough to generate a stable value for $I(\mathbf{V}, \mathbf{x}; \mathbf{r}|\mathbf{H})$. Note here that MI, which is evaluated using (5), always saturates at $\mathcal{H}(\mathbf{V}, \mathbf{x})$, which is defined as [34]

$$\mathcal{H}(\mathbf{V}, \mathbf{x}) = \mathcal{E}_{\mathbf{V}, \mathbf{x}} \{-\log_2 (\Pr(\mathbf{V}, \mathbf{x}))\} \quad (7)$$

as $E_s/\sigma^2 \rightarrow \infty$, where $\mathcal{H}(\cdot)$ denotes the differential entropy.

III. CURRENT CAPACITY RESULT AND ITS DRAWBACKS

The existing analysis calculates the capacity in two stages. If the channel equation in (1) is considered, the MI between the channel input and the output can be written as

$$I(\mathbf{x}, \mathbf{V}; \mathbf{r}) = I(\mathbf{x}; \mathbf{r}|\mathbf{V}) + I(\mathbf{V}; \mathbf{r}) \quad (8)$$

where $I(\mathbf{x}; \mathbf{r}|\mathbf{V})$ is the MI between the constellation symbol and $\mathbf{r}$, and $I(\mathbf{V}; \mathbf{r})$ is the MI achieved by modulating information in the antenna index. The $I(\mathbf{x}; \mathbf{r}|\mathbf{V})$ is evaluated by assuming the optimum input distribution for $\mathbf{x}$, denoted here as $f_{\mathbf{x}}$, is a Gaussian variate, and the corresponding capacity is

$$C_1 = \max_{f_{\mathbf{x}}} I(\mathbf{x}; \mathbf{r}|\mathbf{V}) = \frac{1}{L} \sum_{i=1}^L \log_2 \left| 1 + \frac{pE_s}{T\sigma^2} \tilde{\mathbf{H}}_i \tilde{\mathbf{H}}_i^H \right|. \quad (\text{A})$$

The second term on the right-hand side of (8) is more complex. It is evaluated in the existing analysis by presuming the channel input distribution $f_{\mathbf{x}}$ as Gaussian; thus, the channel transfer probability $\Pr(\mathbf{r}|\mathbf{V} = \tilde{\mathbf{H}}_i)$ is

$$\Pr(\mathbf{r}|\mathbf{V} = \tilde{\mathbf{H}}_i) = \frac{1}{\pi^{n_R} |\Sigma|} e^{-\mathbf{r}^H \Sigma^{-1} \mathbf{r}} \quad (9)$$

where $\Sigma = \sigma^2 \mathbf{I} + (pE_s/T) \tilde{\mathbf{H}}_i \tilde{\mathbf{H}}_i^H$, and the *a priori* probability $\Pr(\mathbf{V} = \tilde{\mathbf{H}}_i) = 1/L$. The MI between $\mathbf{V}$ and $\mathbf{r}$ can be written as

$$I(\mathbf{V}; \mathbf{r}) = \frac{1}{L} \sum_{i=1}^L \int \Pr(\mathbf{r}|\mathbf{V} = \tilde{\mathbf{H}}_i) \times \log_2 \left(\frac{\Pr(\mathbf{r}|\mathbf{V} = \tilde{\mathbf{H}}_i)}{\Pr(\mathbf{r})} \right) d\mathbf{r} \quad (10)$$

where $\Pr(\mathbf{r}) = 1/L \sum_{i=1}^L \Pr(\mathbf{r}|\mathbf{V} = \tilde{\mathbf{H}}_i)$. The exact evaluation of (10) seems difficult. From Monte Carlo integration method as in (5), we have

$$I(\mathbf{V}; \mathbf{r}) \approx \frac{1}{Ln} \sum_{i=1}^L \sum_{l=1}^n \log_2 \left(\frac{\Pr(\mathbf{r}^l|\mathbf{V} = \tilde{\mathbf{H}}_i)}{\Pr(\mathbf{r}^l)} \right). \quad (\text{B})$$

The conclusion in [28] is that the sum of C_1 and $I(\mathbf{V}; \mathbf{r})$ in (10) gives the capacity of SM. However, a fundamental weakness of these assumptions is that the input distribution, $f_{\mathbf{x}}$ is Gaussian variate. It appears that $f_{\mathbf{x}}$ affects both $I(\mathbf{x}; \mathbf{r}|\mathbf{V})$ and $I(\mathbf{V}; \mathbf{r})$. It is clear that Gaussian input distribution maximizes $I(\mathbf{x}; \mathbf{r}|\mathbf{V})$, but it is unclear whether it maximizes $I(\mathbf{V}; \mathbf{r})$. The Gaussian assumption simplifies the analysis, but apparently it is not always optimal (see Section IV-C). Further, since Gaussian inputs do not maximize $I(\mathbf{V}; \mathbf{r})$, the upper limit of MI achieved by SM is not clear. It appears that the optimum input distribution for $I(\mathbf{V}; \mathbf{r})$ depends on CSIT, but in the absence of CSIT, $f_{\mathbf{x}}$ cannot be adapted. Furthermore, for instance, when $n_R = 1$ and $T = 1$, $I(\mathbf{V}; \mathbf{r})$ reduces to $I(v; r)$, and it is mainly dependent on how much each channel element differs from other elements. This distinction of elements depends on the underlying fading distribution. In a massive MIMO regime, CSIT is not required to capture the distinctions between channel elements, but the CDI is sufficient. Therefore, in the massive MIMO regime, the actual value of each channel element is not important, but the distribution information is. In this paper, it is shown that CDI in massive MIMO systems can be used to adapt the input distribution to fading to achieve high spectral efficiency. The analysis in this paper presumes no distribution for the channel inputs and provides an upper limit for MI achieved by SM, i.e., SM capacity.

IV. MAIN RESULT AND ANALYSIS

The main result of this paper can be summarized as follows.

Theorem 1: The capacity of the SM MIMO system described by (1) in Section II can be upper bounded by

$$C \leq \log_2 \left| \mathbf{I} + \frac{pE_s}{TL\sigma^2} \mathbf{H} \mathbf{D} \mathbf{H}^H \right| \text{ bits/s/Hz} \quad (11)$$

where the diagonal matrix $\mathbf{D} = \text{diag}(d_1, \dots, d_N)$ depends on the active antenna set, and $\sum_i d_i = TL$. In the limit of $N/n_R \rightarrow \infty$, the capacity upper bound in (11) is almost certainly achievable, and the optimum distribution for $\mathbf{x}$ is independent of CSI but CDI. The channel input distribution should be adapted to the channel fading in order to achieve this upper bound. The optimum per stream input distribution is given by

$$f_{X_i}(x_i) = g(x_{I_i})g(x_{Q_i}) \quad (12)$$

where

$$g(s) = \phi(s) \left(1 + \sum_{\ell=2}^{\infty} A_{2\ell} \text{He}_{2\ell} \left(\sqrt{\frac{2T}{E_s}} s \right) \right) \quad (13)$$

$$\phi(s) = \sqrt{\frac{T}{\pi E_s}} e^{-\frac{T s^2}{E_s}} \quad (14)$$

with $\text{He}_{2\ell}(\cdot)$ is the 2ℓ th probabilists' Hermite polynomial, and

$$A_{2\ell} = \sum_{k=0}^{\ell} \frac{(-1)^k}{k!(2\ell - 2k)! 2^k} m_{2\ell - 2k}^{x_{I_i}}. \quad (15)$$

The ℓ th raw moment of x_{I_i} , i.e., $m_{\ell}^{x_{I_i}}$ is given as

$$m_{\ell}^{x_{I_i}} = \begin{cases} \frac{(\ell-1)!! E_s^{\frac{\ell}{2}} (\mathcal{E}\{h_I^2\})^{\frac{\ell}{2}}}{(2T)^{\frac{\ell}{2}} \mathcal{E}\{h_I^{\ell}\}}, & \text{if } \ell \text{ is even} \\ 0, & \text{if } \ell \text{ is odd.} \end{cases} \quad (16)$$

Note that $m_{\ell}^{x_{I_i}} = m_{\ell}^{x_{Q_i}}$ for all ℓ and the full input distribution is given as $f_{\mathbf{x}} = \prod_i f_{X_i}$.

Theorem 1 shows that the optimum channel input distribution depends on the underlying fading distribution through its raw moments. This is in complete contrast to the conventional SMX MIMO where the optimum channel input is always Gaussian variate. In SMX MIMO, adapting to channel fading often means adapting the input variance (i.e., power) to the fading. The family of the channel input distribution remains unchanged as Gaussian. However, in SM MIMO, adapting to fading means changing the family of distribution. In this context, results from (12)–(16) will be useful. In practice, $\mathbf{x}$ can be designed to have the distribution in (13) using variable-length coding schemes such as Huffman coding. Since (13) is a continuous distribution, it is not possible to design channel inputs exactly to have this continuous distribution, but it can be approximated accurately [31]. The process includes truncation and quantization of the probability distribution. The granularity of the quantization (i.e., Δx_{I_i}) depends on the number of bits to be encoded per real dimension.

A. Proof of Theorem 1

How the matrix, $\mathbf{D}$, is evaluated is given in Example 1. The proof of Theorem 1 is the focus here. Firstly, a simple case is considered, and then, extensions to general antenna cases are followed.

1) $n_R = T = I$: An SM system with a deterministic channel is considered with $n_R = T = 1$. For a single antenna receiver, the channel matrix $\mathbf{H}$ reduces to a vector denoted by $\mathbf{h} = (h_1, \dots, h_N)$, and the channel input–output relationship for SM in (1) can be rewritten as

$$r = \sqrt{p}vx + n \quad (17)$$

where the received signals r, v , and x and AWGN n are now complex scalars. The transmit power is constrained to $\mathcal{E}\{|x|^2\} = E_s$. First, the channel is modeled as a two-source multiplicative multiple access channel (MAC). The sources v and x are independent. It is further assumed that the source v has a finite alphabet that is $h_1, \dots, h_N$. The objective is to calculate the sum capacity of the system. From the chain rule for MI of MAC

$$I(x, v; r) = I(x; r) + I(v; r|x) \quad (18)$$

$$= \mathcal{H}(r) - \mathcal{H}(r|x, v). \quad (19)$$

By following the standards, it can be shown that $\mathcal{H}(r|x, v) = \log_2 \pi e \sigma^2$ in the AWGN channel. Since $I(x, v; r)$ is maximum when $\mathcal{H}(r)$ is maximum, maximizing the entropy of r remains. The $\mathcal{H}(r)$ can be upper bounded by the entropy of a complex Gaussian RV with a variance as same as that of r . Therefore, an upper bound for the capacity of SM can be derived as

$$C = \max_{f_x} I(x, v; r) \leq \log_2 \left(1 + \frac{p\sigma_z^2}{\sigma^2} \right) \quad (20)$$

where σ_z^2 is the second moment of z , and $z = vx$. It is clear from the definition of z that $\sigma_z^2 = \mathcal{E}(|v|^2)E_s$. The capacity problem is now reduced to find the second moment of the source v . Due to v having a finite alphabet, its constellation energy is given by $\sum_{k=1}^N |h_k|^2/N = \|\mathbf{h}\|^2/N$. Note that this is because the transmitter equally likely activates antennas in the absence of CSIT. As a result, a capacity upper bound can be derived as

$$C \leq \log_2 \left(1 + \frac{\rho}{N} \|\mathbf{h}\|^2 \right) \quad (21)$$

where $\rho = pE_s/\sigma^2$. At this stage, it is not clear if the upper bound in (21) could actually be achievable with no CSIT. If the upper bound is to be achieved in a Gaussian noise channel, z should be a Gaussian variate. It appears that CSIT is essential to make z Gaussian variate. Moreover, it is not clear whether z could be made Gaussian variate even with ideal CSIT. However, an interesting scenario occurs in the massive MIMO regime where the capacity upper bound is indeed achievable with no CSIT. The following Lemma provides a useful result for proving the main theorem.

Lemma 1: Let z be a complex RV defined by $z = vx$. Let v be a uniformly distributed complex discrete RV with a support of $v \in \{h_1, \dots, h_N\}$. The elements (i.e., h_k) of the alphabet of v are drawn from a complex random process with finite raw moments. Let $\mathcal{E}(h_I^i)$ and $\mathcal{E}(h_Q^i)$ be the i th raw moments of real and imaginary parts, respectively, of this random process. $x = x_I + jx_Q$ is a zero-mean E_s -variance complex RV with

i.i.d. real and imaginary components. Note that $-\infty \leq h_{Ii}, h_{Qi} \leq \infty$ for all i . In the limit of N (i.e., $N \rightarrow \infty$), z converges in distribution to a Gaussian RV, i.e.,

$$z \xrightarrow{d} \mathcal{CN} \left(0, \frac{\|\mathbf{h}\|^2}{N} E_s \right) \quad (22)$$

where $\xrightarrow{d}$ denotes the convergence in distribution if x varies with the following distribution function:

$$f_X(x) = g(x_I)g(x_Q) \quad (23)$$

where

$$g(s) = \phi(s) \left(1 + \sum_{\ell=2}^{\infty} A_{2\ell} \text{He}_{2\ell} \left(\sqrt{\frac{2}{E_s}} s \right) \right) \quad (24)$$

$$\phi(s) = \frac{1}{\sqrt{\pi E_s}} e^{-\frac{s^2}{E_s}} \quad (25)$$

$$A_{2\ell} = \sum_{k=0}^{\ell} \frac{(-1)^k}{k!(2\ell - 2k)! 2^k} m_{2\ell-2k}^{x_I}. \quad (26)$$

The i th raw moment of x_I , i.e., $m_{\ell}^{x_I}$

$$m_{\ell}^{x_I} = \begin{cases} \frac{(\ell-1)!! E_s^{\frac{\ell}{2}} (\mathcal{E}\{h_I^{\ell}\})^{\frac{\ell}{2}}}{2^{\frac{\ell}{2}} \mathcal{E}\{h_I^{\ell}\}}, & \text{if } \ell \text{ is even} \\ 0, & \text{if } \ell \text{ is odd.} \end{cases} \quad (27)$$

Note that $m_{\ell}^{x_I} = m_{\ell}^{x_Q}$ for all ℓ .

Proof: See Appendix B. ■

Lemma 1 shows that in massive MIMO scenarios (i.e., $N \rightarrow \infty$), an input distribution can be found for x , which makes z a Gaussian variate, and it does not depend on v . However, it further shows that the input distribution depends on CDI. Unlike in single-input–single-output fading channels, the capacity achieving x is not always Gaussian. Therefore, in the absence of CSIT or strictly with no CSIT but with CDIT, the capacity upper bound in (21) is indeed achievable as $N \rightarrow \infty$. Hence, the capacity of SM with a deterministic wireless channel in a massive MIMO regime can be given as

$$C = \log_2 \left(1 + \frac{\rho}{N} \|\mathbf{h}\|^2 \right) \quad (28)$$

and the capacity in fading C_F becomes

$$C_F = \mathcal{E}_{\mathbf{h}} \left\{ \log_2 \left(1 + \frac{\rho}{N} \|\mathbf{h}\|^2 \right) \right\} \quad (29)$$

where $\mathcal{E}_{\mathbf{h}}\{\cdot\}$ denotes the expectation over fading coefficients. The expectation in (29) can be evaluated for many fading models accurately. For instance, in Rayleigh fading, $\|\mathbf{h}\|^2$ is distributed as

$$f_{\|\mathbf{h}\|^2}(t) = \frac{1}{\Gamma(N)} t^{N-1} e^{-t}, \quad \text{for } t \geq 0. \quad (30)$$

Hence, the ergodic capacity in Rayleigh fading can be evaluated as

$$C_F = \int_0^{\infty} \log_2 \left(1 + \frac{\rho t}{N} \right) f_{\|\mathbf{h}\|^2}(t) dt. \quad (31)$$

The integral in (31) can be solved in closed form and is given explicitly in [32, pp. 572].

2) $n_R > 1, T = 1$: Next, the results are extended to the $n_R > 1$ case. It can be seen that the source $\mathbf{v}$ is a vector variable. Its discrete alphabet spans from $\mathbf{h}_1$ to $\mathbf{h}_N$. The capacity upper bound can be calculated for a fixed $\mathbf{H}$ as

$$C = \max_{f_x} I(x, \mathbf{v}; r) = \log_2 \left| \mathbf{I} + \frac{1}{\sigma_z^2} \mathbf{Q}_z \right| \quad (32)$$

where the vector variable $\mathbf{z} = \mathbf{v}x$, and $\mathbf{Q}_z$ is the covariance matrix of $\mathbf{z}$. $|\cdot|$ denotes the matrix determinant. It is clear that if $\mathbf{z}$ is to be Gaussian distributed, its elements (i.e., z_i for $i = 1, \dots, n_R$) should be Gaussian variate, but $z_i = v_i x$, where v_i is the i th element of $\mathbf{v}$, is a complex scalar product. Therefore, the input distribution derived in Lemma 1 for x makes each element of $\mathbf{z}$ a Gaussian variate, making the vector $\mathbf{z}$ jointly Gaussian variate. From the independence of x and $\mathbf{v}$, it is obtained that

$$\mathbf{Q}_z = \mathcal{E}_{\mathbf{v}} \{ \mathbf{v} \mathbf{v}^H \} E_s. \quad (33)$$

From the law of large numbers, it can be shown that as $N/n_R \rightarrow \infty$ $\mathcal{E}_{\mathbf{v}} \{ \mathbf{v} \mathbf{v}^H \} = (1/N) \sum_k \mathbf{h}_k \mathbf{h}_k^H = (1/N) \mathbf{H} \mathbf{H}^H$. For example a random experiment is setup in which $\mathbf{v}$ is observed for τ number of trials. Then, the sample average is taken. If τ is sufficiently large, the sample covariance converges to the distribution covariance. Because $\mathbf{h}_k$ are equally likely, for a large but finite number of trials of τ , the observer should obtain $\mathbf{h}_k, \tau/N$ times. Therefore, the sample covariance converges to $(1/N) \mathbf{H} \mathbf{H}^H$. This proves that $(1/N) \mathbf{H} \mathbf{H}^H$ is indeed the covariance of $\mathbf{v}$. Substituting this in (32), the final result is obtained as

$$C = \log_2 \left| \mathbf{I} + \frac{\rho}{N} \mathbf{H} \mathbf{H}^H \right| \quad (34)$$

and the ergodic capacity in fading as

$$C_F = \mathcal{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I} + \frac{\rho}{N} \mathbf{H} \mathbf{H}^H \right| \right\}. \quad (35)$$

3) $n_R = 1, T > 1$: The capacity of SM for a fixed channel depends on the particular selection of antennas sets. The general baseband equation reduces for $n_R = 1$ to

$$r = \sqrt{\rho} \mathbf{g} \mathbf{x} + n \quad (36)$$

where $\mathcal{C}^{1 \times T}$ vector $\mathbf{g} = (g_1, \dots, g_T)$ has a discrete alphabet, the transmit $\mathcal{C}^{T \times 1}$ vector $\mathbf{x} = (x_1, \dots, x_T)^T$ has equal power constellation symbols, and the power of each symbol is set to E_s/T . First, an upper bound is obtained for the sum capacity $I(\mathbf{x}, \mathbf{g}; r)$ as

$$C = \max_{f_{\mathbf{x}}} I(\mathbf{x}, \mathbf{g}; r) \leq \log_2 \left(1 + \frac{\sigma_z^2}{\sigma^2} \right) \quad (37)$$

where σ_z^2 is the second moment of $\tilde{z}$. In this case, $\tilde{z} = \mathbf{g} \mathbf{x}$. Since, $\mathbf{x}$ and $\mathbf{g}$ are independent, it can be obtained that

$$\begin{aligned} \sigma_z^2 &= \mathcal{E}_{\mathbf{g}} \{ \mathbf{g} \mathbf{g}^H \} \frac{E_s}{T} \\ &= \mathcal{E}_{g_1, \dots, g_T} \{ g_1 g_1^* + \dots + g_T g_T^* \} \frac{E_s}{T}. \end{aligned} \quad (38)$$

If transmit antenna k is activated d_k times in total, σ_z^2 reduces to (see Example 1)

$$\sigma_z^2 = \frac{E_s}{LT} \mathbf{h}^H \mathbf{D} \mathbf{h} \quad (39)$$

where the diagonal matrix $\mathbf{D} = \text{diag}(d_1, \dots, d_N)$. Substituting the result in (39) into (37), the desired result for the upper bound is obtained. The alphabets for g_1 to g_T depend on the selection of the transmit antenna sets. The alphabet of each g_i has a different number of distinct elements, and the same element may occur multiple times. The following is defined to capture these phenomena. Let N_i ($N_i \leq N$) be the number of distinct elements in the alphabet of g_i . Let the alphabet of g_i be $g_i \in \{e_{i1}, \dots, e_{iN_i}\}$, and assume that the e_{ik} th element occurs a_{ik} times in the alphabet of g_i , and all a_{ik} 's are integers. Note that although a different notation is introduced for the alphabet of g_i , here, they are actually drawn from the original channel matrix $\mathbf{H}$ and, thus, from the random process Φ , as shown in Example 1. Hence, the ℓ th raw moments of real and imaginary parts of e_{ik} are $\mathcal{E}\{h_I^\ell\}$ and $\mathcal{E}\{h_Q^\ell\}$, respectively, $\forall i, k$. Further, due to the constraint on the total number of active antenna sets, $\sum_k a_{ik} = L$ for all i . To achieve the capacity upper bound in (37) in a Gaussian noise channel, $\tilde{z}$ should be made a Gaussian variate. $\tilde{z}$ is a sum of complex scalar products: $\tilde{z} = \sum_i g_i x_i$. If each product can be made Gaussian, as a result, $\tilde{z}$ becomes Gaussian. Hence, an arbitrary product denoted as $g_i x_i$ is considered, and the per stream input distribution is denoted as f_{X_i} . Let e_{ik} be $e_{ik} = e_{Iik} + j e_{Qik}$, where e_{Iik} is the real part of e_{ik} . The per dimension input distribution of $x_i = x_{Ii} + j x_{Qi}$ is considered here. We then have the characteristic function (CF) from (93) for x_{Ii} as

$$\varphi_{x_{Ii}}(t) = \sum_{\ell=0}^{\infty} m_{\ell}^{x_{Ii}} \frac{(jt)^\ell}{\ell!} \quad (40)$$

where

$$m_{\ell}^{x_{Ii}} = \begin{cases} \frac{(\ell-1)!! E_s^{\frac{\ell}{2}} \left(\sum_{k=1}^{N_i} a_{ik} e_{Iik}^2 \right)^{\frac{\ell}{2}}}{T^{\frac{\ell}{2}} L^{\frac{\ell-2}{2}} \left(\sum_{k=1}^{N_i} a_{ik} e_{Iik}^\ell \right)}, & \text{if } \ell \text{ is even} \\ 0, & \text{if } \ell \text{ is odd.} \end{cases} \quad (41)$$

From Lemma 5, in the limit of $N_i \rightarrow \infty$, regardless of the weights a_{ik}

$$\frac{1}{L} \left(\sum_{k=1}^{N_i} a_{ik} e_{Iik}^\ell \right) = \mathcal{E}\{h_I^\ell\} \quad (42)$$

if $\sum_{k=1}^{N_i} a_{ik}^2 / (\sum_{k=1}^{N_i} a_{ik})^2 \rightarrow 0$. Since all a_{ik} 's are integers, this condition is always satisfied, but depending on the actual values of a_{ik} , the convergence speed may be changed. The convergence speed could be improved by selecting active antenna sets, which result $N_i \approx N$. Then, (41) becomes

$$m_{\ell}^{x_{Ii}} = \begin{cases} \frac{(\ell-1)!! E_s^{\frac{\ell}{2}} (\mathcal{E}\{h_I^2\})^{\frac{\ell}{2}}}{(2T)^{\frac{\ell}{2}} \mathcal{E}\{h_I^\ell\}}, & \text{if } \ell \text{ is even} \\ 0, & \text{if } \ell \text{ is odd.} \end{cases} \quad (43)$$

This confirms that the per dimension input distribution does not depend on CSI but CDI. Similar to the approach in Lemma 1, the per stream per dimension input distribution denoted $f_{X_{Ii}}$ can be derived, and the full per stream input distribution f_{X_i} can be obtained as in Theorem 1. Due to the channel model considered in this paper (see Section II), alphabets of all g_i 's are drawn from the same random process. As a result, all x_i 's are identically distributed. The independent stream transmission implies $f_{\mathbf{x}} = \prod_i f_{X_i}$. Therefore, it is clear that, in the massive MIMO regime, $\tilde{\mathbf{z}}$ can be made a Gaussian variate with no CSIT. Substituting the result in (39) into (37), the desired result is obtained for the capacity of SM with a deterministic wireless channel in the massive MIMO regime, i.e.,

$$C = \log_2 \left(1 + \frac{\rho}{LT} \mathbf{h}^H \mathbf{D} \mathbf{h} \right) \quad (44)$$

where $\sum_{k=1}^N d_k = LT$, and the capacity in fading is

$$C_F = \mathcal{E}_{\mathbf{h}} \left\{ \log_2 \left(1 + \frac{\rho}{LT} \mathbf{h}^H \mathbf{D} \mathbf{h} \right) \right\}. \quad (45)$$

The following example further explains the analysis.

Example 1: Consider an SM system where the number of transmit antennas is set to $N = 4$ and the number of active antennas is set to $T = 2$ for exposition. Let the allowed antenna sets be (1, 2), (1, 3), (1, 4), (2, 3), making $L = 4$. The alphabet for v_1 is h_1, h_2 and for v_2 is h_2, h_3, h_4 ; thus, $N_1 = 2$ and $N_2 = 3$. Therefore

$$\mathcal{E}_{v_1} \{v_1 v_1^H\} = \frac{1}{L} (3|h_1|^2 + |h_2|^2) \quad (46)$$

$$\mathcal{E}_{v_2} \{v_2 v_2^H\} = \frac{1}{L} (|h_2|^2 + 2|h_3|^2 + |h_4|^2). \quad (47)$$

After substituting (46) and (47) into (38), and with some simplifications, we have $\sigma_{\tilde{\mathbf{z}}}^2 = (E_s/8) \mathbf{h}^H \mathbf{D} \mathbf{h}$, where in this particular case, $\mathbf{D} = \text{diag}(3, 2, 2, 1)$. An SM system with the modified antenna set (1, 2), (3, 1), (4, 3), (2, 4) may be more favorable from an information-theoretic perspective because $N = N_1 = N_2 = 4$, i.e., the maximum allowed by the system dimension.

4) $n_R > 1$, $T > 1$: The analysis follows the same logic as for a single-antenna case. The channel equation is given in (1). The details are not given here, and the final result for a fixed channel matrix $\mathbf{H}$ is given by

$$C = \log_2 \left| \mathbf{I} + \frac{\rho}{LT} \mathbf{H} \mathbf{D} \mathbf{H}^H \right| \quad (48)$$

and the ergodic capacity in fading is

$$C_F = \mathcal{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I} + \frac{\rho}{LT} \mathbf{H} \mathbf{D} \mathbf{H}^H \right| \right\}. \quad (49)$$

One may compare (49) with the capacity results in the absence of CSIT in [2]. In general, the diagonal matrix $\mathbf{D}$ may have duplicate entries. Hence, the evaluation of the expectation in (49) is quite complex but can be evaluated exactly as shown in [33].

B. Special Case

In Rayleigh fading, samples of $\mathbf{V}$ are drawn from a zero mean complex Gaussian random process, and we have the following result for the density function of $\mathbf{x}$.

Corollary 1: In Rayleigh fading and in the limit of $N \rightarrow \infty$, the capacity achieving per stream input distribution f_{X_i} converges in distribution to

$$f_{X_i}(x_i) = g(x_I)g(x_Q) \quad (50)$$

where the per dimension density function is given by

$$g(s) = \frac{1}{2} \delta \left(s + \sqrt{\frac{E_s}{2T}} \right) + \frac{1}{2} \delta \left(s - \sqrt{\frac{E_s}{2T}} \right) \quad (51)$$

where $\delta(s)$ is the direct delta function.

Proof: Here, the per dimension density function of X_i is derived, and the results of Theorem 1 is applied. From (16), in Rayleigh fading

$$\mathcal{E} \{h_I^2\} = \frac{1}{2}, \text{ and } \mathcal{E} \{h_I^\ell\} = (\ell - 1)!! \left(\frac{1}{2} \right)^{\frac{\ell}{2}}. \quad (52)$$

Application of (52) into (16) directly yields that $m_\ell^{x_{Ii}} = (E_s/2T)^{\ell/2}$ for ℓ is even and zero otherwise. It gives the CF of x_{Ii}

$$\begin{aligned} \varphi_{x_{Ii}}(t) &= 1 - \frac{E_s t^2}{2T \cdot 2!} + \frac{E_s^2 t^4}{4T^2 \cdot 4!} - \frac{E_s^3 t^6}{8T^3 \cdot 6!} + \dots \\ &= \sum_{\ell=0}^{\infty} \frac{(-1)^\ell (\sqrt{E_s} t)^{2\ell}}{(2T)^\ell \cdot 2\ell!} = \cos \left(\sqrt{\frac{E_s}{2T}} t \right). \end{aligned} \quad (53)$$

By inverting CF for x_{Ii} in (53), the optimum per dimension input distribution in Rayleigh fading can be found as shown in (51). ■

Similarly, an input distribution can be obtained for any fading scenario such as Rician and Nakagami fading. Moreover, this approach could be employed to derive the optimal input distribution for channels that do not have a statistical fading model but measured data. In this context, (16) may be particularly useful, where sample moments may be used instead of distribution moments. For illustrative purposes, the results of a Monte Carlo simulation are given in Fig. 2 to confirm the result presented in Corollary 1.

C. Remarks

The most favorable fading scenario for SM in a Gaussian noise channel is Rayleigh fading. In undesirable fading scenarios where different antenna sets are difficult to distinguish at the receiver, the information transferred in the antenna set index is degraded. In SM, the total capacity depends on the individual capacities achieved by both constellation symbols (e.g., $C_{\mathbf{x}}$) and antenna set index (e.g., $C_{\mathbf{V}}$). In undesirable fading, SM capacity is mainly governed by $C_{\mathbf{x}}$. For instance, if all channel elements are approximately equal, $C_{\mathbf{V}} \rightarrow 0$; thus, SM capacity reduces to the capacity of a $n_R \times T$ MIMO system, where $f_{\mathbf{x}}$ converges to a T -variate jointly Gaussian distribution because $\mathcal{E} \{h_I^\ell\} = (\mathcal{E} \{h_I^2\})^{\ell/2} \forall \ell$ in (16). Therefore, channel

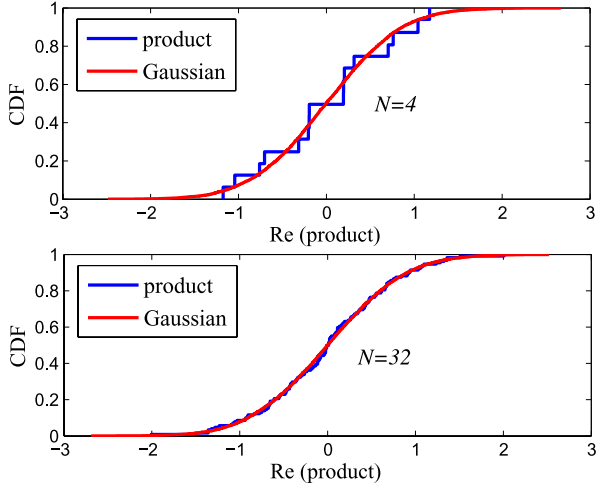


Fig. 2. Cumulative distribution function comparison of the real part of the product vx in (17) for different N in Rayleigh fading. Optimum input distribution in (50) is used, where $E_s = 0$ dB, and $T = 1$.

input distribution should be adjusted in accordance with the fading scenario. In this context, the analysis in this paper can be useful to derive the optimum channel input distribution for a given fading scenario; thus, the dependence of the input distribution on the distribution information of fading is not surprising. In Rician fading for instance, the optimum channel input distribution varies with the Rician K factor. An SM system in an i.i.d. Rician fading environment is considered with $N = 64$, $T = 1$, and $n_R = 2$. The channel matrix $\mathbf{H}$ is modeled as [41]

$$\mathbf{H} = \sqrt{\frac{K}{1+K}} \mathbf{H}_d + \sqrt{\frac{1}{1+K}} \mathbf{H}_s \quad (54)$$

where $\mathbf{H}_d$ and $\mathbf{H}_s$ are the deterministic and specular component matrices of $\mathbf{H}$, respectively. The matrix $\mathbf{H}_s$ contains elements $h_{ik}^{[s]} \sim \mathcal{CN}(0, 1)$, and the matrix $\mathbf{H}_d$ contains constant elements $h_{ik}^{[d]} = (1/\sqrt{2})(1+j)$. Strictly, $h_{ik}^{[d]}$ is equal to $e^{-j2\pi r_{ik}/\lambda}$, where r_{ik} is the distance between k th transmit antenna to the i th receive antenna, and λ is the carrier wavelength [42]. For transmit and receive antenna arrays with colocated elements, assuming an arbitrary unit norm complex value for $h_{ik}^{[d]}$ can be shown to be statistically equivalent to the more strict model. Hence, the assumption of $h_{ik}^{[d]} = (1/\sqrt{2})(1+j)$ is justified. As a result, the ℓ th raw moment of Φ , i.e., $\mathcal{E}\{h_I^\ell\}$, is given as

$$\mathcal{E}\{h_I^\ell\} = \left(\frac{1}{1+K}\right)^{\frac{\ell}{2}} \mathcal{E}_U \left\{ \left(\sqrt{\frac{K}{2}} + \frac{U}{\sqrt{2}} \right)^\ell \right\} \quad (55)$$

where $U \sim \mathcal{N}(0, 1)$. For instance

$$\mathcal{E}\{h_I\} = \sqrt{\frac{K}{2(1+K)}}, \quad \mathcal{E}\{h_I^2\} = \frac{1}{2} \quad (56a)$$

$$\mathcal{E}\{h_I^3\} = \frac{(K+3)\sqrt{K}}{2\sqrt{2}(1+K)^{\frac{3}{2}}} \quad (56b)$$

$$\mathcal{E}\{h_I^4\} = \frac{K^2 + 6K + 3}{4(1+K)^2} \quad (56c)$$

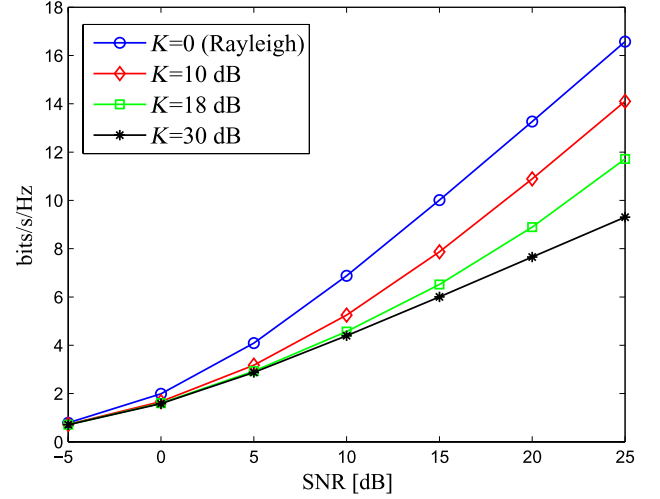


Fig. 3. Ergodic capacity of a 2×64 SM MIMO link in Rician fading.

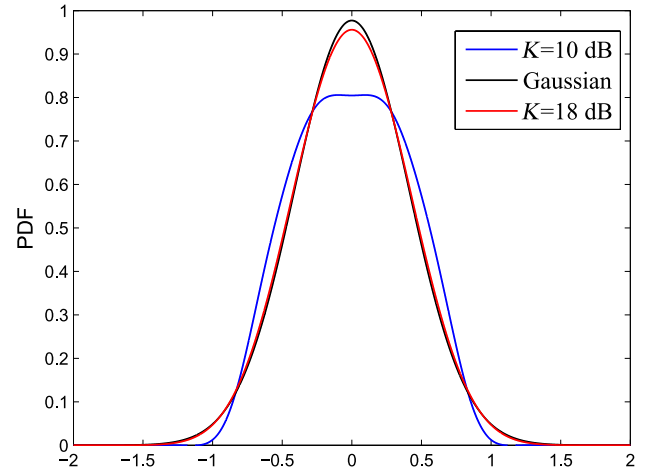


Fig. 4. Per stream per dimension input distribution $f_{X_{Ti}}$ for a SM system in Rician fading, and $E_s = 0$ dB.

are the first four raw moments. The results here in conjunction with (16) confirm that the per stream channel input distribution in Rician fading varies with the K factor. The channel capacity also varies adversely with the K factor. Fig. 3 shows the ergodic capacity [see (49)] of an SM MIMO link in Rician fading, where different K factors are considered. It is apparent that the capacity reduces with an increasing K factor. The system has a maximum of two degrees of freedom (DoF), but when $K = 30$ [dB], it clearly reduces to one. The DoF is the capacity growth rate as defined in [34]. The per dimension per stream input distribution, $f_{X_{Ti}}$ is evaluated using Theorem 1, and results are shown in Fig. 4. The optimum input distribution when $K = 10$ [dB] is clearly different from Gaussian distribution. However, in strong line-of-sight scenarios, e.g., when $K = 18$ [dB], the optimum input distribution approximately coincides with a Gaussian distribution as shown in Fig. 4 (red curve). It is important to note that high K factors are used to merely highlight the fact that, in extremely undesirable fading conditions, Gaussian inputs can also be approximately optimal, but such high K factors may not be common in practice. Therefore, although Gaussian

codebooks are, in general, not optimal for SM, in some undesirable fading scenarios, they can approximately be optimal. The Gaussian codebooks are optimal only in the limit of $K \rightarrow \infty$, which in practice is not relevant. For instance, the moment expression of per stream per dimension distribution of $\mathbf{x}$ in (16) is considered. The fourth-order raw moment of x_{Ii} is

$$m_4^{x_{Ii}} = \frac{3!!E_s^2 (\mathcal{E}\{h_I^2\})^2}{(2T)^2 \mathcal{E}\{h_I^4\}} = \frac{3E_s^2(1+K)^2}{(2T)^2(K^2+6K+3)} \quad (57)$$

but

$$\lim_{K \rightarrow \infty} \frac{(1+K)^2}{K^2+6K+3} = 1. \quad (58)$$

Therefore, $\lim_{K \rightarrow \infty} m_4^{x_{Ii}} = 3E_s^2/4T^2$. Similarly, all even moments converge as

$$\lim_{K \rightarrow \infty} m_{2k}^{x_{Ii}} = \frac{(2k-1)!!E_s^k}{(2T)^k}. \quad (59)$$

Therefore, the CF of $x_{Ii}, \varphi_{x_{Ii}}(t)$ is

$$\varphi_{x_{Ii}}(t) = \sum_{k=0}^{\infty} \frac{(-1)^k (2k-1)!! (\sqrt{E_s}t)^{2k}}{(2T)^k \cdot 2k!} \quad (60)$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k E_s^k t^{2k}}{(4T)^k \cdot k!} = e^{-\frac{E_s t^2}{4T}}. \quad (61)$$

The inversion of (61) yields $x_{Ii} \sim \mathcal{N}(0, E_s/2T)$ in the limit of $K \rightarrow \infty$. For the purpose of completeness, CDI for several widely used fading models are summarized in Table I.

V. MASSIVE BUT FEW ACTIVE MULTIPLE-INPUT–MULTIPLE-OUTPUT IN CORRELATED RAYLEIGH FADING

Here, the analysis is extended to a scenario where the transmit-side spatial correlation exists. If the transmit and receive antennas are collocated in a small area, the spatial correlation is unavoidable. The effect of receive correlation upon the SM capacity can be readily captured by incorporating receive correlation into (49). It can be shown that the receive correlation has no effect on the channel input distribution. It appears that the transmit-side correlation in contract could have a significant impact on the input distribution, and this effect is investigated here. Without loss of generality, an SM system with $n_R = T = 1$ is considered. From Theorem 1, the per stream per dimension input distribution of x depends on

$$m_\ell^{x_I} = \begin{cases} \frac{(\ell-1)!!E_s^{\frac{\ell}{2}} (\mathcal{E}\{h_I^2\})^{\frac{\ell}{2}}}{2^{\frac{\ell}{2}} \mathcal{E}\{h_I^\ell\}}, & \text{if } \ell \text{ is even} \\ 0, & \text{if } \ell \text{ is odd.} \end{cases} \quad (62)$$

Here

$$\mathcal{E}\{h_I^\ell\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N h_{Ii}^\ell, \quad \ell = 1, 2, \dots \quad (63)$$

but the channel vector $\mathbf{h} = (h_1, \dots, h_N)$ in correlated Rayleigh fading scenarios has correlated elements. From [35]–[37],

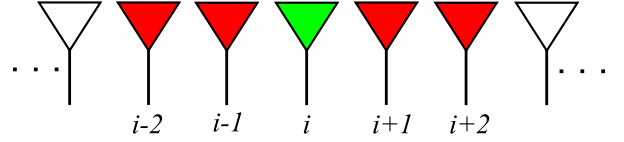


Fig. 5. Uniform linear array (ULA). The space correlation occurs from antennas in red to the antenna in green.

the channel vector can be rewritten as

$$\mathbf{h} = \mathbf{h}_w \mathbf{R}_T^{\frac{1}{2}} \quad (64)$$

where $\mathbf{R}_T$ is the transmit-side space correlation matrix, and $\mathbf{h}_w \sim \mathcal{CN}(0, \mathbf{I})$. In this model, if $n_R \geq 1$, $\mathbf{H} = \mathbf{H}_w \mathbf{R}_T^{1/2}$. Here, the correlation between two transmit antennas is restricted to be the same irrespective of the receive antenna where it is observed. It is shown in [38] that this constraint is often satisfied by space diversity arrays. According to the model in (64), it can be seen that each and every transmit antenna is correlated more or less with all other transmit antennas. This is somewhat a strong assumption. As shown in the Fig. 5, the green (the i th) antenna is correlated only with four close-by antennas. The correlation occurs from other antennas are insignificant and thus negligible. If it is assumed that the area occupied by the large-scale MIMO system is scaled with N such that the spatial separation between antenna elements is fixed, and the channel elements $\{h_i\}$ can be modeled as an m -dependent random sequence (see Definition 2). It is $m = 2$ in the scenario shown in Fig. 5 and $m = 0$ for uncorrelated fading. Typically, each antenna is correlated with neighboring antennas in correlated fading. For instance, the i th antenna could be correlated with $i - 2$ th, $i - 1$ th, $i + 1$ th and $i + 2$ th antennas, as shown in Fig. 5. A typical correlation model between the channel elements can be given as [35]

$$\mathcal{E}\{h_{Ii}h_{Ii'}\} = \begin{cases} 0.5e^{-0.05d^2(i-i')^2}, & |i-i'| \leq m \\ 0, & \text{otherwise} \end{cases} \quad (65)$$

where d is the antenna spacing in wavelengths. Here, the interest is in the quantity; $I_\ell = 1/N \sum_{i=1}^N h_{Ii}^\ell$. From Theorem 3 in Appendix D, it is shown that, in the limit of $N \rightarrow \infty$, I_ℓ converges to a nonrandom quantity that depends only on the CDI, which in this case is the correlation coefficient. Theorem 3 is applied here; let $\{X_i\} = \{h_{Ii}^\ell\}$ be a stationary m -dependent sequence of real RVs. Let $\mathcal{E}\{h_{Ii}^\ell\} = \mu$, $\text{Var}\{h_{Ii}^\ell\} \leq \infty$, and $I_\ell = 1/N \sum_{i=1}^N h_{Ii}^\ell$ be the partial sum. Then, from Theorem 3 in Appendix D, it can be shown that

$$\lim_{N \rightarrow \infty} I_\ell \xrightarrow{d} \mathcal{N}\left(\mu, \frac{\tau^2}{N}\right) \quad (66)$$

where $\tau^2 = \text{Var}\{h_{Ii}^\ell\} + 2 \sum_{l=1}^m \text{Cov}\{h_{Ii}^\ell, h_{Ii+l}^\ell\}$. Since $\{h_{Ii}^\ell\}$ is a m -dependent sequence, τ^2 is bounded. Hence, in the limit of $N \rightarrow \infty$, the standard deviation of the limiting distribution approaches to zero. Therefore, I_ℓ almost certainly converges to μ

$$\lim_{N \rightarrow \infty} I_\ell \rightarrow \mu. \quad (67)$$

From Lemma 6, h_{Ii} can be given as $h_{Ii} = \mathbf{a}^T \tilde{\mathbf{u}}$, where the $R^{m+1 \times 1}$ vector $\tilde{\mathbf{u}} = \mathcal{N}(0, \mathbf{I})$. Here, in normalized correlated Rayleigh fading, $\mathbf{a}^T \mathbf{a} = 1/2$. Then

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N h_{Ii}^\ell = \mu = \mathcal{E}_{\tilde{\mathbf{u}}} \left\{ (\tilde{\mathbf{u}}^T \mathbf{a} \mathbf{a}^T \tilde{\mathbf{u}})^{\frac{\ell}{2}} \right\} \quad (68)$$

$$= (\ell - 1)!! (\mathbf{a}^T \mathbf{a})^{\frac{\ell}{2}} \quad (69)$$

$$= (\ell - 1)!! \left(\frac{1}{2} \right)^{\frac{\ell}{2}} \quad (70)$$

for $\ell = 2, 4, 6, \dots$. The result in (70) in conjunction with Corollary 1 yields the optimal input distribution for SM in correlated fading. It appears that the optimum input distribution in m -correlated Rayleigh fading is the same as that in uncorrelated Rayleigh fading (see Corollary 1), but it may be different in other fading scenarios. The analytical method presented here will also be useful to obtain the optimum input distribution in other fading scenarios.

A. Remarks

The result in Section V is a direct consequence of the limit result in (66). This limit result holds only if the ratio τ^2/N approaches to zero as N goes to infinity. In the m -dependent case, τ^2 is bounded so τ^2/N approaches to zero almost certainly, but in the fully correlated case ($m \rightarrow N$), this limit result also might hold. The parameter τ^2 increases with m , but if the increase is slower than the increase in N , τ^2/N can still approach to zero. In some instances, if the correlation between antennas decays sufficiently fast over the space, τ^2 may converge to a constant. In correlated Rayleigh fading, the convergence of I_ℓ to μ , thus, the input distribution to the theoretical optimum distribution, might be significantly slower, and the convergence speed mainly depends on τ^2 and, hence, on the correlation structure of the antenna array.

VI. NUMERICAL RESULTS AND DISCUSSION

Here, a simulation study for comparing theoretical results obtained in Section IV with the MI attainable with practical constellations in fading is presented. The MI simulations are based on the Monte Carlo method outlined in Section II-A. The MI in fading can be calculated as

$$C_F^{\text{SIM}} = \mathcal{E}_{\mathbf{H}} \{I(\mathbf{V}, \mathbf{x}; \mathbf{r}|\mathbf{H})\}. \quad (71)$$

The receive signal-to-noise ratio (SNR) is set to pE_s/σ^2 . The modulation rate γ (in bits per channel use) denotes the number of bits modulated into the channel in a single channel use. The modulation rate described here is different from the rate widely used in coding literature [14]. Note that, in fact, γ is the channel input entropy in bits. Hence, it is also equivalent to $\gamma = \mathcal{H}(\mathbf{V}, \mathbf{x})$, and it includes both information and redundant bits. Simulation results are from several scenarios as follows.

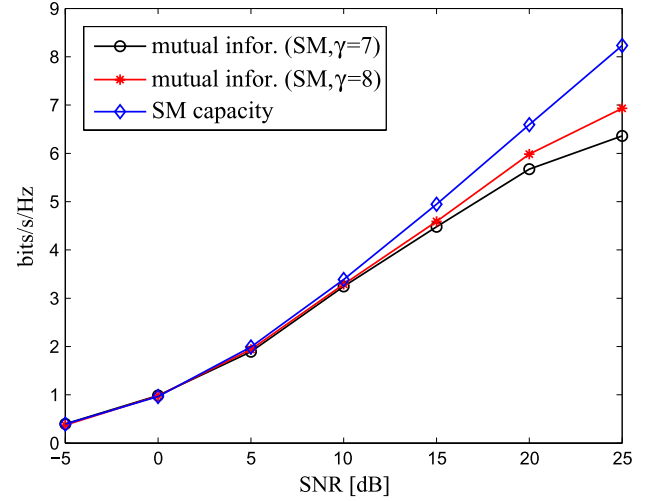


Fig. 6. Ergodic capacity and MI performance of a 1×64 SM MIMO system. Here, $T = 1$, and BPSK/QPSK signal constellations are used for MI evaluation.

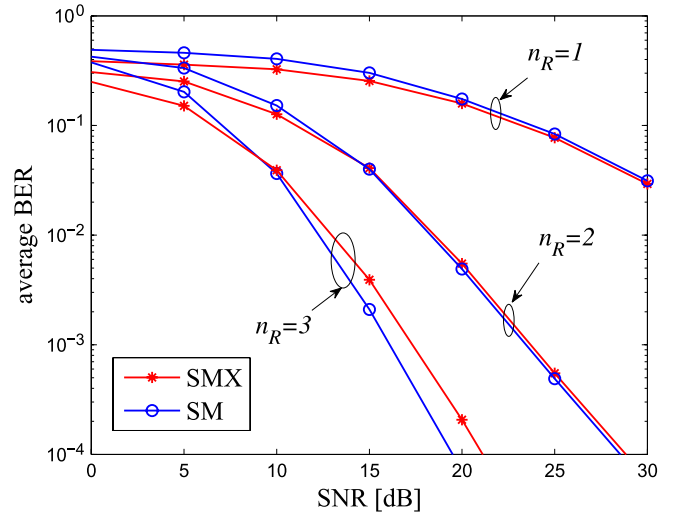


Fig. 7. Uncoded average bit error probability of SM and SMX with $\gamma = 7$, $N = 64$, and $n_R = 1/2/3$.

1) *S1*: The objective of this scenario is to check if the capacity result in (29) is attainable with practical constellations such as QAM/PSK. The results are shown in Fig. 6, where MI curves are shown for binary phase-shift keying (BPSK) and quadrature PSK (QPSK) with 6 bits modulated in the antenna index, i.e., $\mathcal{H}(\mathbf{V}) = 6$. There are a total of 7 and 8 bits modulated in every time slot (i.e., channel use) in BPSK and QPSK constellations, respectively. The MI curves are generated using (5). It is apparent that simulated curves follow the analytical capacity closely until SNR = 10 dB. After that, MI curves start to drift away from the capacity as SNR increases, but the drift is larger in the simulation curve with fewer modulated bits (i.e., when $\gamma = 7$). Further, MI curves saturate at 7/8 bits/s/Hz because the channel input entropy $\mathcal{H}(\mathbf{V}, \mathbf{x})$ is 7 and 8 bits/s/Hz in BPSK and QPSK, respectively.

2) *S2*: In Fig. 7, the uncoded average bit error probability of an SM system with $N = 64$ and $n_R = 1/2/3$ is considered.

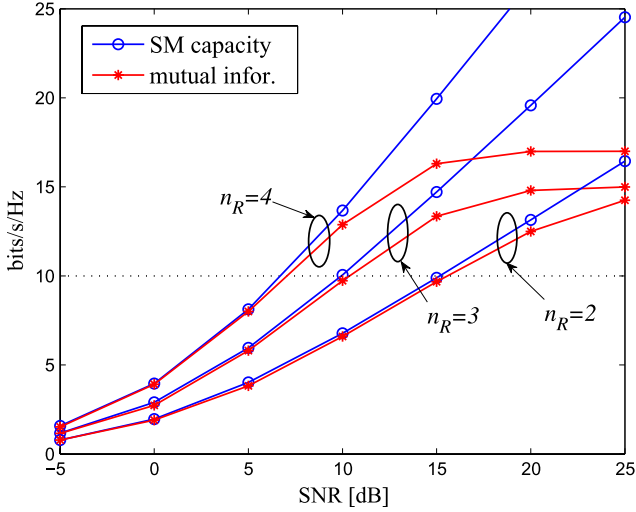


Fig. 8. Ergodic capacity and MI performance of $2/3 \times 64$ SM MIMO systems for $\gamma = 15/17$ in uncorrelated Rayleigh fading.

The error performance of SM in different fading scenarios has been extensively studied [18], [19]. For comparison, an SMX system with seven independent BPSK-modulated signal streams is also considered, where random precoding is used to map seven data streams to 64 transmit antennas [26]. Note that the random beamforming may not be strictly comparable with SM but could be used to obtain baseline error performance in the absence of CSIT. It is shown in Fig. 7 that SM and SMX achieve comparable average bit error rate (BER) at all the SNRs considered. Here, SM uses a BPSK constellation and $T = 1$; thus, $\mathcal{H}(\mathbf{V}) = 6$ and $\mathcal{H}(\mathbf{x}) = 1$.

3) S3: An SM system with $n_R = 2, 3$, and 4 is considered in uncorrelated Rayleigh fading. The ergodic capacity is compared with the achievable MI with $\gamma = 15/17$. Here, $N = 64$, $L = 2^{12}/2^{14}$, and $T = 3$. Accordingly, a $2^{12}/2^{14}$ set of three antenna sets is selected for transmission, where antenna sets are organized as in Example 1 to make $N_i \approx N \forall i$. To achieve $\gamma = 15$, 12 bits are modulated in $L = 2^{12}$ antenna sets, i.e., $\mathcal{H}(\mathbf{V}) = 12$; three BPSK symbols are transmitted directly from T antennas, i.e., $\mathcal{H}(\mathbf{x}) = 3$; and 14 bits are modulated in $L = 2^{14}$ antenna sets and three BPSK symbols are transmitted directly from T antennas for $\gamma = 17$. The results are shown in Fig. 8. The theoretical capacity in (49) is also shown for comparison. The theoretical capacity of SM and the MI with practical constellation are comparable. Moreover, the MI follows the theoretical capacity in SNR as high as 20 dB with $n_R = 2$, but it reduces to 10 dB in $n_R = 3$. The MI curves saturate at 15/17 bits/s/Hz because the channel input entropy $\mathcal{H}(\mathbf{V}, \mathbf{x}) = 15/17$ bits, respectively, for $\gamma = 15/17$.

4) S4: In Fig. 9, the uncoded average bit error probability of a SM system in Rayleigh fading with $N = 64$ is considered. For comparison, a SMX system with 15 independent BPSK modulated signal streams is also considered, where random precoding is used to map 15 data streams to 64 transmit antennas, and $n_R = 3$. It is shown in Fig. 9 that SM and SMX achieve comparable average uncoded BER at all SNRs considered.

5) S5: An SM system with $N = 64$, $n_R = 2, 3$, and $T = 2$ in correlated Rayleigh fading is considered. Here, the channel

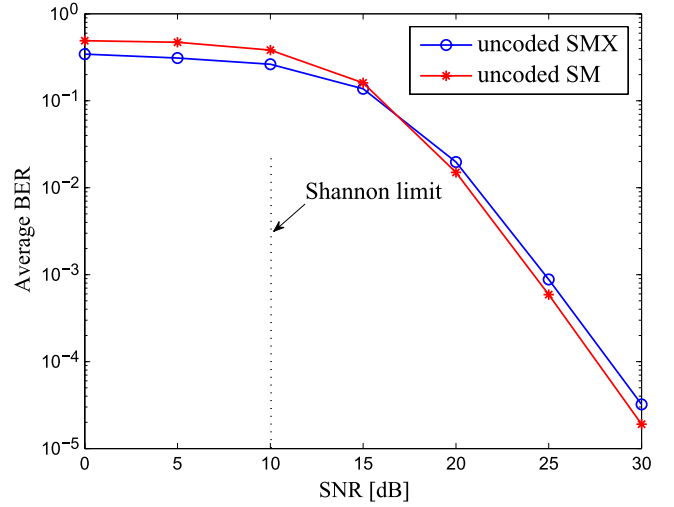


Fig. 9. Uncoded average bit error probability of SM and SMX with $\gamma = 15$, $N = 64$, $n_R = 3$, and $T = 3$.

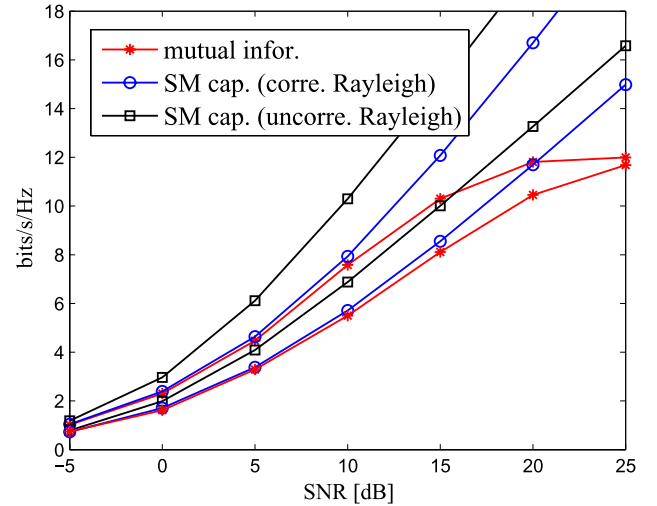


Fig. 10. Ergodic capacity and MI of $2/3 \times 64$ SM MIMO system for $\gamma = 12$ in correlated Rayleigh fading.

matrix $\mathbf{H}$ is modeled as

$$\mathbf{H} = \mathbf{R}_R^{\frac{1}{2}} \mathbf{H}_w \mathbf{R}_T^{\frac{1}{2}} \quad (72)$$

where $\mathbf{R}_R$ and $\mathbf{R}_T$ are receive and transmit correlation matrices, respectively. The correlation matrices are generated using the exponential decay model in [43], where the i, k th element of either $\mathbf{R}_R$ and $\mathbf{R}_T$ is given by

$$[\mathbf{R}_X]_{ik} = r_c^{|i-k|} \quad (73)$$

where $r_c = e^{-\beta}$, with β being the correlation decay coefficient. The subscript X denotes either R or T in (73). A comparison between the SM capacity and the MI achieved by BPSK modulation is shown in Fig. 10. Here, it is assumed that $r_c = 0.8$ at both the receiver and the transmitter. As shown, the capacity is reduced considerably due to the spatial correlation. The MI in both cases follows the capacity curves until the saturation at 12 bits/s/Hz. This is because, in this simulation, $\gamma = 12$ ($\mathcal{H}(\mathbf{V}) = 10$, $\mathcal{H}(\mathbf{x}) = 2$).

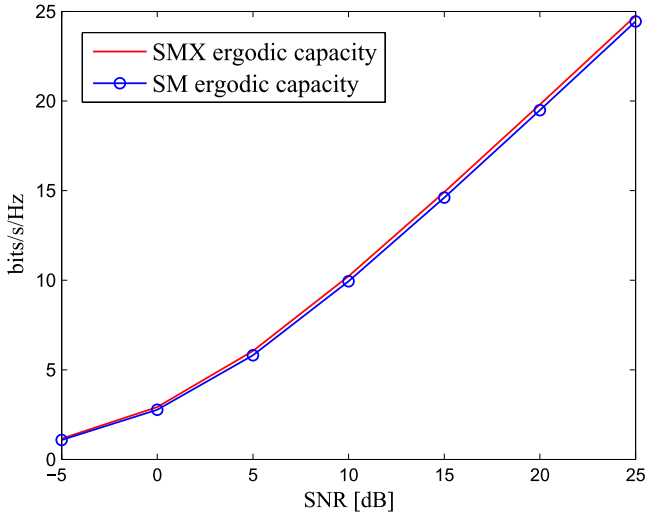


Fig. 11. Ergodic capacity comparison of SM and SMX MIMO in Rayleigh fading, where $N = 64$, and $n_R = 3$.

A. Further Remarks

This paper shows that SM, which is an antenna switching technique, can achieve capacity comparable with the maximum theoretical capacity of open-loop MIMO systems. The ergodic capacity of open-loop SMX MIMO is given by [2]

$$C_{\text{SMX}} = \mathcal{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I} + \frac{gE_s}{N\sigma^2} \mathbf{H}\mathbf{H}^H \right| \right\} \text{ bits/s/Hz.} \quad (74)$$

Referring to the MI curve for $n_R = 3$ in Fig. 8, for instance, a 3×64 MIMO system with SM could theoretically achieve nearly 10 bit/s/Hz of spectral efficiency with a $R = 2/3$ channel code rate at SNR = 10 dB [39]. The capacity curve [i.e., (11)] for $n_R = 3$ in Fig. 8 confirms that 10 bit/s/Hz is very close to the SM capacity. In fact results in Fig. 11 confirms that this is also very close to the open-loop SMX MIMO capacity. This suggests that a strong forward error correction (FEC) coding scheme will ensure that SM in practice achieves rates comparable to the open-loop SMX MIMO. Finding the optimum FEC coding scheme is outside the scope of this paper, but for example, the uncoded BER and Shannon limit shown in Fig. 9 defines the playing field for channel coding for this running example [40].

Theorem 1 quantifies the SM capacity and proves the achievability with no CSIT. It further quantifies the optimum input distribution as $N/n_R \rightarrow \infty$. However, it does not provide how the SM should be adapted to fading for finite N . SM modulation should be carefully designed for different fading scenarios for finite N to achieve (or approximate) the SM capacity predicted in Theorem 1. Here, example SM modulation schemes are considered in different fading conditions. Let there be an SM system with $N = 64$ and $n_R = 2$ in Rayleigh and Rician fading. The number of active antennas is set to $T = 2$, and the input entropies are set to $\mathcal{H}(\mathbf{V}) = 10$ and $\mathcal{H}(\mathbf{x}) = 2$, resulting $\gamma = 12$. Here, $\mathcal{H}(\mathbf{x}) = 2$ is achieved by using two independent BPSK data streams. As shown in Fig. 12, the achievable MI in Rayleigh fading is closely comparable to SM capacity until about 15 dB. In this case, 2^{10} active antenna sets are

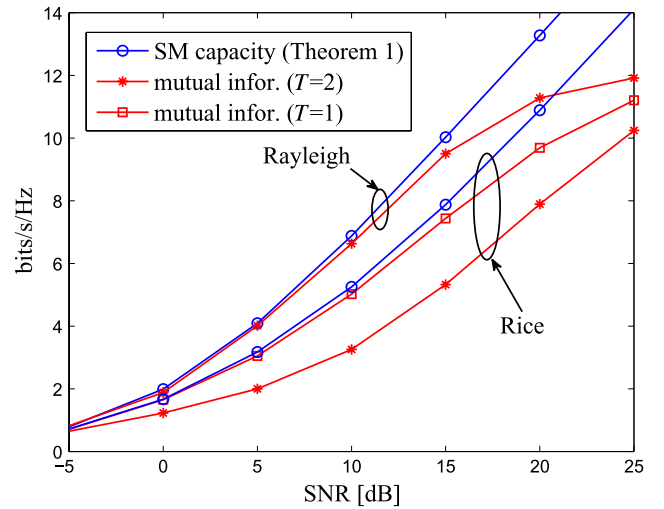


Fig. 12. Ergodic capacity and MI performance comparison of SM in Rayleigh and Rician fading. Here, $K = 10$ dB for Rician fading.

selected in such a way that $\mathbf{D}$ is approximately a scaled identity matrix. However, in Rician fading, achievable MI is dramatically reduced. The reason is that the SM modulator modulates 10 bits out of 12 bits in the antenna index, and the fading is relatively undesirable to support such a high rate in the antenna index. Hence, the sum MI (i.e., the sum of MI achieved by the antenna index and the constellation symbol) is not close to the SM capacity predicted by Theorem 1. One can conclude that SM capacity cannot be achieved in this scenario. However, if $\mathcal{H}(\mathbf{V})$ is reduced to 6 bits, and $\mathcal{H}(\mathbf{x})$ is increased to 6 bits, it can be seen that achievable MI denoted by square markers is again closely comparable to the SM capacity in Rician fading. Here, $\mathcal{H}(\mathbf{V}) = 6$ is achieved by setting $T = 1$, and $\mathcal{H}(\mathbf{x}) = 6$ is achieved by using a single 64-QAM data stream. Therefore, in SM, modulation should be carefully adapted to fading conditions. It may include adaptation of the channel input distribution $f_{\mathbf{x}}$ and/or bit allocation to spatial and constellation dimensions, i.e., $\mathcal{H}(\mathbf{V})$ and $\mathcal{H}(\mathbf{x})$.

There is another interesting scenario that occurs for finite N and $n_R > T$. The number of bits encoded in the antenna index is inevitably limited to $\mathcal{H}(\mathbf{V}) \leq \log_2 \lfloor \frac{N}{T} \rfloor$. Hence, if L antenna sets are considered, $I(\mathbf{V}; \mathbf{r})$ in (B) converges to $\log_2 L$ as SNR = E_s/σ^2 goes to infinity. The sum MI is simply dominated by MI achieved by the constellation symbols, i.e., $I(\mathbf{x}; \mathbf{r}|\mathbf{V})$. The growth rate of $I(\mathbf{x}; \mathbf{r}|\mathbf{V})$ as $E_s/\sigma^2 \rightarrow \infty$ can at most be $\min(n_R, T)$, which in this case is T . One can conclude that SM cannot achieve n_R DoF in this case, but this behavior can be attributed to the fact that N is finite. In this case, typical MI curves exhibit two distinct regions, as shown in Fig. 13. One region that is denoted as the DoF region has n_R DoF, and after a certain SNR, the MI growth rate reduces to T . The region in which SM achieves T DoF is denoted in this paper as the saturation region. The transition SNR can be accurately calculated, and let it be defined as SNR_T . An upper bound for the saturation region can be readily found, and it is given by the summation of $\log_2 L$, and (A). The transition SNR, SNR_T is defined as the SNR at which the high SNR asymptotic for the upper bound for the saturation region and

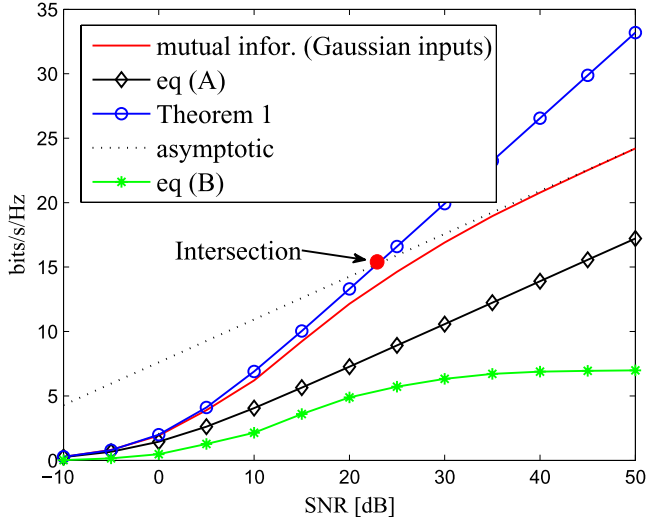


Fig. 13. Capacity and MI performance for finite N and $n_R > T$. Here, it is set that $N = 128$, $T = 1$, and $n_R = 2$.

the SM capacity in (11) intersect, i.e., the red dot in Fig. 13. Any SM system operates below SNR_T for a given N , and T should still have n_R DoFs. Moreover, as $N \rightarrow \infty$, SNR_T goes to infinity. This limitation can easily be reduced or overcome in practice by using as many or more active antennas than the number of receive antennas.

VII. CONCLUSION

In this paper, the information-theoretic capacity limit of SM in the absence of CSI at the transmitter has been analyzed. The particular interest is on systems with a large number of transmit antennas in comparison with the number of receive antennas. It is shown that SM has the potential to achieve the capacity comparable with the open-loop MIMO systems, although a subset of antennas at a time is activated. This is because SM modulates information in the antenna index, which is in complete contrast to the plain antenna switching techniques. Unlike the conventional MIMO operation, both the channel coefficients and input symbols carry information. As a result, SM compensates for the loss of information capacity due to activating a single or a few antennas by modulating information in the antenna index. Then, the sum information capacity remains high. It has been further shown through Monte Carlo simulation for MI that the theoretical limits can be reached with practical constellations such as M -QAM in operationally important SNR regions. It is believed that these results will be useful for energy and spectral efficient wireless data transmission for large-scale MIMO systems in the absence of CSIT.

APPENDIX A SOME USEFUL RESULTS

Here, the complex scalar product, $z = vx$, which is similar to that in (17), is considered. The objective here is to find the probability density function (pdf) of x , which makes z a Gaussian variate when multiplied with v . In the context of this paper, it is the channel input distribution. Towards that objective, we have the following definitions and results.

Definition 1: Let $X = X_I + jX_Q$ be a random complex scalar RV; X_I and X_Q be its real and imaginary components, respectively; and $\sqrt{-1} = j$. Let the marginal pdfs of X_I be g_{X_I} and of X_Q be g_{X_Q} . The joint pdf of X_I and X_Q is $g_{X_I X_Q}$, and the pdf of complex scalar X f_X is defined as

$$f_X(x) = g_{X_I X_Q}(x_I, x_Q). \quad (75)$$

If real and imaginary components are independent, the pdf of X is given by $f_X = g_{X_I}(x_I)g_{X_Q}(x_Q)$. If marginal pdfs are independently and identically distributed, the pdf of X is completely characterized by a marginal pdf of either X_I or X_Q . In that scenario, the marginal pdf of X_I or X_Q is denoted the per dimension pdf of X .

Lemma 2: [44] Let $\{F_n\}$ denotes a sequence of distribution functions with finite moments of all orders. Let the k th moment of the distribution $\{F_n\}$ be denoted by

$$\beta_{n,k} = \int x^k dF_n(x). \quad (76)$$

If for each fixed integer $k \geq 0$ $\beta_{n,k}$ converges to a finite limit β_k , the sequence of distribution functions $\{F_n\}$ converge to a limit F , which has the same moment sequence $\{\beta_k\}$.

Lemma 3: [44] Let F be a measure (distribution function) on the real axis such that all the moments

$$\beta_k = \int x^k dF(x), \quad k = 0, 1, 2, \dots \quad (77)$$

are finite. If

$$\sum_{n=1}^{\infty} \beta_{2n}^{-\frac{1}{2n}} = +\infty \quad (78)$$

F is the only measure on the real axis with $\{\beta_k\}$ as its sequence of moments. This is also known as Carleman's condition in classical analysis.

Lemma 4: Let F be a measure (distribution function) on the real axis such that all the moments:

$$\beta_k = \int x^k dF(x), \quad k = 0, 1, 2, \dots \quad (79)$$

are finite. Since F is a probability measure $\beta_0 = 1$. The pdf of F , i.e., f can be approximated by

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\beta_1)^2}{2\beta_2}} \sum_{i=0}^{\infty} A_i \text{He}_i \left(\frac{t-\beta_1}{\sqrt{\beta_2}} \right) \quad (80)$$

where $\text{He}_i(t)$ is the i th probabilists' Hermite polynomial, and

$$A_i = \frac{1}{i!} \int_{-\infty}^{\infty} \text{He}_i(t) f(t) dt. \quad (81)$$

The n th probabilists' Hermite polynomial is

$$\text{He}_n(t) = n! \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m}{m!(n-2m)! 2^m} t^{n-2m} \quad (82)$$

for $-\infty \leq t \leq \infty$. This approximation is also known as the Gram-Charlier series.

Theorem 2: Let V be a uniformly distributed real discrete random variable with support of $V \in \{v_1, \dots, v_N\}$, and X is a real zero-mean E_x -variance random variable. The elements (i.e., v_k) of the alphabet of V are drawn from a random process, i.e., $\mathcal{V}$ with finite raw moments. The i th raw moment of the random process is denoted by $\mathcal{E}(v^i)$. Note that $-\infty \leq v_i \leq \infty$ for all i . Let the real random variable Z be given by

$$Z = VX. \quad (83)$$

Let g_X be the pdf of X . In the limit of $N \rightarrow \infty$

$$Z \xrightarrow{d} \mathcal{N}\left(0, \frac{\left(\sum_{k=1}^N v_k^2\right)}{N} E_x\right) \quad (84)$$

if X varies with the following density function:

$$g_X(s) = \phi(s) \left(1 + \sum_{i=2}^{\infty} A_{2i} \text{He}_{2i}\left(\frac{s}{\sqrt{E_x}}\right)\right) \quad (85)$$

where

$$\phi(s) = \frac{1}{\sqrt{2\pi E_x}} e^{-\frac{s^2}{2E_x}} \quad (86)$$

$$A_{2i} = \sum_{k=0}^i \frac{(-1)^k}{k!(2i-2k)!2^k} m_{2i-2k}^X. \quad (87)$$

The i th raw moment of X , m_i^X

$$m_i^X = \begin{cases} \frac{(i-1)!! E_x^{\frac{i}{2}} (\mathcal{E}\{v^2\})^{\frac{i}{2}}}{\mathcal{E}\{v^i\}}, & \text{if } i \text{ is even} \\ 0, & \text{if } i \text{ is odd.} \end{cases} \quad (88)$$

The RV X converges in distribution to an independent random variable that is independent of V but dependent on the distribution information of V .

Proof: If it is assumed that V and X are independent for the time being, we have

$$m_i^Z = \frac{\left(\sum_{k=1}^N v_k^i\right)}{N} m_i^X \quad (89)$$

where m_i^Z and m_i^X are the i th raw moments of Z and X , respectively. The second moment of Z is given by

$$m_2^Z = \frac{\left(\sum_{k=1}^N v_k^2\right)}{N} E_x \quad (90)$$

but from Gaussian assumption for Z , we know

$$m_i^Z = \begin{cases} (m_2^Z)^{\frac{i}{2}} (i-1)!!, & \text{if } i \text{ is even} \\ 0, & \text{if } i \text{ is odd} \end{cases} \quad (91)$$

where $(i-1)!!$ denotes the double factorial, i.e., the product of every odd numbers from $i-1$ to 1. From (89)–(91), we have

$$m_i^X = \begin{cases} \frac{(i-1)!! E_x^{\frac{i}{2}} \left(\sum_{k=1}^N v_k^2\right)^{\frac{i}{2}}}{N^{\frac{i-2}{2}} \left(\sum_{k=1}^N v_k^i\right)}, & \text{if } i \text{ is even} \\ 0, & \text{if } i \text{ is odd.} \end{cases} \quad (92)$$

It can be verified that if m_i^X satisfies the moment equation for X in (92), in conjunction with (89), the moments of Z satisfy the Gaussian requirements [i.e., (91)]. Therefore, it confirms that if X has the aforementioned structure for its raw moments, Z is certainly Gaussian distributed: $Z \sim \mathcal{N}(0, (\sum_{k=1}^N v_k^2/N) E_x)$. We then have the CF for X as

$$\varphi_X(t) = \sum_{i=0}^{\infty} m_i^X \frac{(jt)^i}{i!} \quad (93)$$

but for a stationary random process

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N v_k^i = \mathcal{E}\{v^i\} \quad i = 1, 2, \dots \quad (94)$$

Therefore, for each fixed integer $i \geq 0$, m_i^X converges to

$$m_i^X = \begin{cases} \frac{(i-1)!! E_x^{\frac{i}{2}} (\mathcal{E}\{v^2\})^{\frac{i}{2}}}{\mathcal{E}\{v^i\}}, & \text{if } i \text{ is even} \\ 0, & \text{if } i \text{ is odd.} \end{cases} \quad (95)$$

From moments convergence theorem in Lemma 2, it can be shown that X has a limit distribution. Then, the pdf of X , i.e., g_X , can be derived from

$$\begin{aligned} g_X(s) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(t) e^{-jst} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{i=0}^{\infty} m_i^X \frac{(jt)^i}{i!} e^{-jst} dt. \end{aligned} \quad (96)$$

It is assumed here that $\varphi_X(t)$ converges. The convergence analysis is possible if m_i^X has an analytical solution; otherwise, the Carleman's condition in Lemma 3 should be used to verify the existence of a unique measure. An analytical expression for m_i^X may not exist if the underlying random process $\mathcal{V}$ does not have an analytical distribution but a finite number of samples. The inversion in (96) can be evaluated approximately, and the resultant approximation (expansion) about a normal distribution is well documented and is generally termed as the Gram–Charlier series (82). From (80), we have

$$g_X(s) = \phi(s) \left(1 + \sum_{i=2}^{\infty} A_{2i} \text{He}_{2i}\left(\frac{s}{\sqrt{E_x}}\right)\right) \quad (97)$$

where

$$\phi(s) = \frac{1}{\sqrt{2\pi E_x}} e^{-\frac{s^2}{2E_x}} \quad (98)$$

$$A_{2i} = \sum_{k=0}^i \frac{(-1)^k}{k!(2i-2k)!2^k} m_{2i-2k}^X. \quad (99)$$

Here, A_2 and all A s with odd indexes can be shown to be zero using the fact that all odd raw moments of X are zero. Further, from (95), it can be verified that, when $N \rightarrow \infty$, the X converges in distribution to an independent random variable that is independent of V . However, from (95), it is further clear

that the distribution, g_X , depends on the moments of the fading distribution and, thus on the distribution information of V . ■

APPENDIX B PROOF OF LEMMA 1

Consider the complex scalar product of $z = vx$. Let v and x be given as $v = v_I + jv_Q$, and $x = x_I + jx_Q$. Then, z can be expanded as

$$z = v_I x_I - v_Q x_Q + jv_I x_Q + jv_Q x_I. \quad (100)$$

If z is to be Gaussian distributed, from (100), each product in (100) should be a Gaussian variate. The alphabets of v_I and v_Q are the real and imaginary parts of $\{h_1, \dots, h_N\}$. For instance, $v_I \in \{h_{I1}, \dots, h_{IN}\}$. Therefore, without loss of generality, the first product on the right-hand side of (100) is considered. It is a product of a real discrete random variable and a real continuous random variable. From Theorem 2, it is clear that the first product $v_I x_I$ can be made a Gaussian variate, i.e., $v_I x_I \sim \mathcal{N}(0, (\sum_{k=1}^N h_{Ik}^2 / 2N) E_s)$, if the pdf of x_I is

$$g_{x_I}(s) = \phi(s) \left(1 + \sum_{i=2}^{\infty} A_{2i} \text{He}_{2i} \left(\sqrt{\frac{2}{E_s}} s \right) \right) \quad (101)$$

where

$$\phi(s) = \frac{1}{\sqrt{\pi E_s}} e^{-\frac{s^2}{E_s}}. \quad (102)$$

The constant, A_{2i} , is given in (26). The fact that $\mathcal{E}\{x_I^2\} = E_s/2$ is used. Since v_I and v_Q are identically distributed, x_I makes the first and the last product in the right-hand side of (100) a Gaussian variate. Similarly, a distribution function, $g_{x_Q}(s)$, for x_Q , which is identical to $g_{x_I}(s)$, will also make the second and third products of (100) Gaussian variates with corresponding variances. Since $x = x_I + jx_Q$, the complex input distribution with zero mean and E_s variance can be obtained from Definition 1 and (101), i.e.,

$$f_X(x) = g(x_I)g(x_Q) \quad (103)$$

where the resulting distribution, i.e., the sum of four Gaussian RVs of z can be found, and it is given in Lemma 1.

APPENDIX C PROOF OF LEMMA 5

Lemma 5: Let $\{x_i\}$ be a sequence of i.i.d. RVs with finite mean and variance drawn from a random process, and let $\{a_i\}$ be a sequence of positive weights; define $A_N = \sum_{i=1}^N a_i$ and $J_k = (1/A_N) \sum_{i=1}^N a_i x_i^k$. Let $\mathcal{E}\{x^k\}$ be the k th raw moment of the random process. If the weight sequence $\{a_i\}$ satisfies the following condition:

$$\lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N a_i^2}{\left(\sum_{i=1}^N a_i\right)^2} = 0 \quad (104)$$

in the limit of $N \rightarrow \infty$, regardless of the weights $J_k \rightarrow \mathcal{E}\{x^k\}$.

Proof: The mean and variance of J_k are considered as follows:

$$\mathcal{E}\{J_k\} = \frac{1}{A_N} \sum_{i=1}^N a_i \mathcal{E}\{x_i^k\} = \mathcal{E}\{x^k\} \quad (105)$$

$$\begin{aligned} \text{Var}\{J_k\} &= \mathcal{E} \left\{ \frac{1}{A_N^2} \left(\sum_{i=1}^N a_i x_i^k \right) \left(\sum_{i=1}^N a_i x_i^k \right) \right\} - \mathcal{E}\{x^k\}^2 \\ &= \mathcal{E} \left\{ \frac{1}{A_N^2} \left(\sum_{i=1}^N a_i^2 x_i^{2k} + \sum_{i \neq l} a_i a_l x_i^k x_l^k \right) \right\} - \mathcal{E}\{x^k\}^2 \\ &= \frac{\sum_{i=1}^N a_i^2}{A_N^2} \mathcal{E}\{x^{2k}\} + \frac{\sum_{i \neq l} a_i a_l}{A_N^2} \mathcal{E}\{x^k\}^2 - \mathcal{E}\{x^k\}^2 \\ &= \frac{\sum_{i=1}^N a_i^2}{A_N^2} \left(\mathcal{E}\{x^{2k}\} - \mathcal{E}\{x^k\}^2 \right) \\ &= \frac{\sum_{i=1}^N a_i^2}{\left(\sum_{i=1}^N a_i\right)^2} \text{Var}\{x^k\}. \end{aligned} \quad (106)$$

From (106), it is clear that if $\sum_{i=1}^N a_i^2 / \left(\sum_{i=1}^N a_i\right)^2 \rightarrow 0$, then $\text{Var}\{J_k\} \rightarrow 0$; therefore, $J_k \rightarrow \mathcal{E}\{x^k\}$. This completes the proof. ■

APPENDIX D PRELIMINARIES FOR SECTION V

Definition 2: [45], [46] $X_1, X_2, \dots$ is a sequence of real random variables with $\mathcal{E}\{X_i\} = \mu$ and $\mathcal{E}\{X_i^2\} \leq \infty$. The sequence of random variables $\{X_i\}$ is said to be m -dependent if $|s - t| \geq m$, random variable, X_t is independent of X_s . In other words, random variables with indexes t and s that are more than m apart are independent.

Lemma 6: If $\{X_i\}$ is a m -dependent Gaussian random sequence, and $\{U_i\}$ be an uncorrelated Gaussian random sequence, X_i could be given as a linear combination of uncorrelated Gaussians random variables in $\{U_i\}$. For instance, X_i can be given as

$$X_i = a_1 U_i + a_2 U_{i-1} \quad (107)$$

if $\{X_i\}$ is a 1-dependent sequence. Moreover, X_i can be given as

$$X_i = a_1 U_{i-1} + a_2 U_i + a_3 U_{i+1} \quad (108)$$

if $\{X_i\}$ is a 2-dependent random sequence. The coefficients a_i 's in the right-hand sides of (107) and (108) depend on the correlation structure of the original sequence $\{X_i\}$.

Theorem 3: [45] Let $\{X_i\}$ be a stationary m -dependent sequence of real random variables. Let $\mathcal{E}\{X_i\} = \mu$, $\text{Var}\{X_i\} \leq \infty$, and $S_n = 1/n \sum_{i=1}^n X_i$ be the partial sum. Then

$$\lim_{n \rightarrow \infty} \sqrt{n}(S_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma_m^2) \quad (109)$$

where $\sigma_m^2 = \text{Var}\{X_k\} + 2 \sum_{i=1}^m \text{Cov}\{X_k, X_{k+i}\}$. Here, variance is $\text{Var}\{X_k\} = \mathcal{E}\{X_k^2\} - \mu^2$, and the covariance is $\text{Cov}\{X_k, X_{k+i}\} = \mathcal{E}\{X_k X_{k+i}\} - \mu^2$.

ACKNOWLEDGMENT

D. A. Basnayaka gratefully acknowledges the insightful discussions with Prof. Peter. J. Smith at University of Canterbury, Christchurch, New Zealand.

REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [2] E. I. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
- [3] D. Gesbert, M. Kountouris, R. W. Heath, C. B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.
- [4] D. A. Basnayaka, P. J. Smith, and P. A. Martin, "Performance analysis of macrodiversity MIMO systems with MMSE and ZF receivers in flat rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2240–2251, May 2013.
- [5] E. Biglieri *et al.*, *MIMO Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
- [6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [7] L. Lu *et al.*, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 11, pp. 742–758, Apr. 2014.
- [8] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
- [9] D. J. Love *et al.*, "An overview of limited feedback wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.
- [10] W. Webb, *Wireless Communications: The Future*. New York, NY, USA: Wiley, 2007.
- [11] X. Chu *et al.*, *Heterogeneous Cellular Networks*. Cambridge, U.K.: Cambridge Univ. Press, May 2013.
- [12] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [13] G. J. Foschini, "Layered spacetime architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Syst. Tech. J.*, vol. 1, pp. 41–59, 1996.
- [14] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [15] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Tech.*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.
- [16] M. Di Renzo, H. Haas, A. Ghayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities and implementation," *Proc. IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.
- [17] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalised spatial modulation," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1498–1502.
- [18] J. Jeganathan, A. Ghayeb, and L. Szczecinski, "Spatial modulation: Optimal detection and performance analysis," *IEEE Commun. Lett.*, vol. 12, no. 8, pp. 545–547, Aug. 2008.
- [19] M. Di Renzo and H. Haas, "Bit error probability of SM-MIMO over generalized fading channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1124–1144, Mar. 2012.
- [20] E. Basar, U. Aygolu, E. Panayirci, and H. Poor, "Performance of spatial modulation in the presence of channel estimation errors," *IEEE Commun. Lett.*, vol. 16, no. 2, pp. 176–179, Feb. 2012.
- [21] K. Ishibashi and S. Sugiura, "Effects of antenna switching on band-limited spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 345–348, Aug. 2014.
- [22] X. Wu, H. Claussen, M. Di Renzo, and H. Haas, "Channel estimation for spatial modulation," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4362–4372, Dec. 2014.
- [23] S. Sugiura and L. Hanzo, "Effects of channel estimation on spatial modulation," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 805–808, Dec. 2012.
- [24] A. Stavridis, S. Sinanovic, M. Di Renzo, and H. Haas, "Energy evaluation of spatial modulation at a multi-antenna base station," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2013, pp. 1–5.
- [25] K. Takeuchi, "Spatial modulation achieves information-theoretically optimal energy efficiency," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1133–1136, May 2015.
- [26] M. Kountouris, D. Gesbert, and T. Salzer, "Enhanced multiuser random beamforming: Dealing with the not so large number of users case," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1536–1545, Oct. 2008.
- [27] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
- [28] Y. Yang and B. Jiao, "Information-guided channel-hopping for high data rate wireless communication," *IEEE Commun. Lett.*, vol. 12, no. 4, pp. 225–227, Apr. 2008.
- [29] X. Guan, Y. Xiao, Y. Cai, and W. Yang, "On the mutual information and precoding for spatial modulation with finite alphabet," *IEEE Wireless Commun. Lett.*, vol. 2, no. 4, pp. 383–386, Aug. 2013.
- [30] D. A. Basnayaka and H. Haas, "Spatial modulation for massive MIMO," in *Proc. IEEE ICC*, May 2015, pp. 1945–1950.
- [31] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
- [32] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Boston, MA, USA: Academic, 2007.
- [33] M. Chiani, M. Z. Win, and H. Shin, "MIMO networks: The effects of interference," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 336–349, Jan. 2010.
- [34] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [35] T.-S. Chu and L. J. Greenstein, "A semi-empirical representation of antenna diversity gain at cellular and PCS base stations," *IEEE Trans. Commun.*, vol. 45, no. 6, pp. 644–646, Jun. 1997.
- [36] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
- [37] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Key-holes, correlations and capacities of multi-element transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 361–368, Apr. 2002.
- [38] A. Moustakas, H. Baranger, L. Balents, A. Sengupta, and S. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, Jan. 2000.
- [39] B. M. Hochwald *et al.*, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2013.
- [40] D. J. Costello and G. D. Forney, Jr., "Channel coding: The road to channel capacity," *Proc. IEEE*, vol. 95, no. 6, pp. 1150–1177, Jun. 2007.
- [41] S. K. Jayaweera and H. V. Poor, "On the capacity of multi-antenna systems in the presence of Rician fading," in *Proc. IEEE Veh. Technol. Conf.*, Vancouver, BC, Canada, Sep. 2002, pp. 1963–1967.
- [42] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [43] A. V. Zelst and J. S. Hammerschmidt, "A single coefficient spatial correlation model for multiple-input multiple-output (MIMO) radio channels," in *Proc. 27th Gen. Assembly URSI*, Maastricht, The Netherlands, Aug. 2002, pp. 1–4.
- [44] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. New York, NY, USA: Springer-Verlag, 2010.
- [45] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, vol. 15, pp. 773–780, 1948.
- [46] P. H. Diananda, "The central limit theorem for m-dependent variables," in *Proc. Cambridge Philos. Soc.*, 1955, vol. 51, pp. 92–95.



Dushyantha A. Basnayaka (S'11–M'12) received the B.Sc.Eng. degree (first-class honors) from the University of Peradeniya, Peradeniya, Sri Lanka, in 2006 and the Ph.D. degree in electrical engineering from the University of Canterbury, Christchurch, New Zealand, in 2012.

From 2006 to 2009, he was with MillenniumIT (a member company of the London Stock Exchange Group). From 2009 to 2012, he was with the Communication Research Group, University of Canterbury. He is currently with the Institute for

Digital Communications, The University of Edinburgh, Edinburgh, U.K. He is a holder of one U.S. patent. His current research interests include massive multiple-input multiple-output, spatial modulation, interference mitigation techniques for cellular wireless systems, coordinated multipoint systems, and visible-light communication systems.

Dr. Basnayaka received the University of Canterbury International Doctoral Scholarship for his doctoral studies and the Best Paper Award at the Spring 2015 IEEE Vehicular Technology Conference (IEEE VTC-Spring 2015).



Marco Di Renzo (S'05–A'07–M'09) was born in L'Aquila, Italy, in 1978. He received the Laurea (*cum laude*) and Ph.D. degrees in electrical and information engineering from the Department of Electrical and Information Engineering, University of L'Aquila, in April 2003 and January 2007, respectively.

From August 2002 to January 2008, he was with the Center of Excellence for Research DEWS, University of L'Aquila. From February 2008 to April 2009, he was a Research Associate with the Telecommunications Technological Center of Catalonia, Barcelona, Spain. From May 2009 to December 2009, he was an Environmental and Physical Science Research Council (EPSRC) Research Fellow with the Institute for Digital Communications, The University of Edinburgh, Edinburgh, U.K. Since January 2010, he has been a Tenured Researcher ("Chargé de Recherche Titulaire") with the French National Center for Scientific Research (CNRS) and a faculty member of the Laboratory of Signals and Systems (L2S), which is a joint research laboratory of the CNRS, the École Supérieure d'Électricité (SUPÉLEC), and the University of Paris-Sud XI, Paris, France. He is currently a Principal Investigator of three European-funded research projects (Marie Curie ITN-GREENET, Marie Curie IAPP-WSN4QoL, and Marie Curie ITN-CROSSFIRE). His main research interests include wireless communications theory.

Dr. Di Renzo currently serves as the Editor for the IEEE COMMUNICATIONS LETTERS. He received special mention for his outstanding five-year academic career (1997–2003) and the THALES Communications Fellowship for doctoral studies (2003–2006) from the University of L'Aquila; the Best Spin-Off Company Award from the Government of Abruzzo Region, Italy, in 2004; the Torres Quevedo Award for research on ultrawideband systems and cooperative localization for wireless networks (2008–2009) from the Ministry of Science and Innovation, Spain; the "Dérégulation pour l'Encadrement de Thèse" from University of Paris-Sud XI, France, in 2010; the 2012 IEEE CAMAD Best Paper Award from the IEEE Communications Society; the 2012 Exemplary Reviewer Award from the IEEE WIRELESS COMMUNICATIONS LETTERS from the IEEE Communications Society; and the 2013 Fall IEEE Vehicular Technology Conference Best Student Paper Award from the IEEE Vehicular Technology Society.



Harald Haas (S'98–A'00–M'03) received the Ph.D. degree from the University of Edinburgh, Edinburgh, U.K., in 2001.

He is currently the Chair of Mobile Communications and the Director of the LiFi Research and Development Center with the University of Edinburgh. He is a Co-founder and the Chief Scientific Officer of pureLiFi Ltd. He first introduced and coined spatial modulation and LiFi. LiFi was listed among the 50 best inventions in TIME Magazine 2011. He is the author of 300 conference and journal papers, including a paper in *Science*. He is a coauthor of the book *Principles of LED Light Communications Towards Networked Li-Fi* (Cambridge, U.K.: Cambridge University Press, 2015). He is a holder of 31 patents and has more than 30 pending patent applications. His main research interests include optical wireless communications, hybrid optical wireless and radio frequency communications, spatial modulation, and interference coordination in wireless networks.

Dr. Haas was an invited speaker at TED Global 2011, and his talk "*Wireless Data from Every Light Bulb*" has been watched online more than 1.6 million times. He gave a second TED Global lecture in 2015 on the use of solar cells as LiFi data detectors and energy harvesters. He coreceived the recent Best Paper Awards at the 2013 Fall IEEE Vehicular Technology Conference (VTC) held in Las Vegas, NV, USA, and at the 2015 Spring IEEE VTC held in Glasgow, U.K., the EURASIP Best Paper Award from the *Journal on Wireless Communications and Networking* in 2015 and the Jack Neubauer Memorial Award from the IEEE Vehicular Technology Society. He received the prestigious Established Career Fellowship from the Engineering and Physical Sciences Research Council (EPSRC) within Information and Communications Technology in the U.K. in 2012 and the Tam Dalyell Prize in 2013, which was awarded by the University of Edinburgh for excellence in engaging the public with science. He was selected by EPSRC as one of the ten Recognising Inspirational Scientists and Engineers (RISE) Leaders in the U.K. in 2014.