Second-Order Optimality of Generalized Spatial Modulation for MIMO Channels With No CSI

Keigo Takeuchi, *Member, IEEE*

***Abstract*—Generalized spatial modulation (GSM) is investigated for Rayleigh block-fading multiple-input multiple-output (MIMO) channels with no channel state information (CSI). Under a condition for coherence time, GSM with quadrature phase-shift keying (QPSK) is proved to achieve a transmission rate that is equal to the capacity of the MIMO channels up to second order in the low signal-to-noise ratio regime, regardless of the number of simultaneously activated transmit antennas and of whether CSI is available at the receiver or not. This implies that QPSK GSM is *second-order* optimal for the MIMO channels in terms of information-theoretical energy efficiency.**

***Index Terms*—Spatial modulation, multiple-input multiple-output (MIMO) systems, no channel state information (CSI), energy efficiency, information theory.**

I. Introduction

SPATIAL modulation (SM) [1]–[3] is an energy-efficient transmission scheme for multi-antenna systems. Since only one transmit antenna is used during one symbol period, SM can be implemented with a single radio-frequency (RF) chain. In spite of low implementation complexity, SM has been proved to be optimal in terms of energy efficiency (EE) for multiple-input single-output (MISO) channels with perfect channel state information (CSI) at the receiver (CSIR) [4], without taking into account circuit power consumption.

In this letter, the previous work [4] is extended to the case of generalized SM (GSM) for multiple-input multiple-output (MIMO) channels with no CSI. In GSM [5], [6], the number of activated antennas is not limited to one in order to increase the transmission rate. The goal of this letter is to elucidate impacts of activating multiple transmit antennas on EE.

It is important to weaken the simplifying assumption of CSIR to the no CSI assumption. The overhead for estimating CSI is non-negligible for GSM, since pilots cannot be sent from all antennas simultaneously. Several schemes [7]–[10] have been proposed to reduce the overhead. However, the performance of GSM is not information-theoretically understood for the no CSI case, whereas it was evaluated in [4], [11] for the CSIR case. This letter analyzes the information-theoretical EE of GSM with no CSI for low signal-to-noise ratio (SNR).

Manuscript received May 20, 2015; revised August 24, 2015; accepted August 28, 2015. Date of publication September 2, 2015; date of current version December 15, 2015. This work was supported in part by the Grant-in-Aid for Young Scientists (A) (MEXT/JSPS KAKENHI Grant Number 26709029), Japan, and by the Grant-in-Aid for Exploratory Research (JSPS KAKENHI Grant Number 15K13987), Japan. The associate editor coordinating the review of this paper and approving it for publication was J. Lee.

The author is with the Department of Communication Engineering and Informatics, University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: ktakeuchi@uec.ac.jp).

Digital Object Identifier 10.1109/LWC.2015.2475742

It is well known that the EE—defined as the channel capacity divided by SNR—is maximized as the SNR tends to zero, so that the EE analysis in the low SNR regime has been considered in the field of information theory [12]–[14]. The main advantage of the low SNR analysis is that non-Gaussian signaling can be evaluated, regardless of whether CSI is available at the receiver or not. In this letter, we follow [14] to propose and analyze *flash* GSM for the no CSI case.

Throughout this letter, a random matrix with independent circularly symmetric complex Gaussian (CSCG) elements with unit variance is called standard complex Gaussian matrix. See [15] for the standard notation in information theory.

II. System Model

A. Rayleigh Block-Fading MIMO Channels

Consider point-to-point MIMO systems with M transmit antennas and N receive antennas over independent and identically distributed (i.i.d.) Rayleigh block-fading channels with a coherence time of T symbol periods [14], [16]. Neither the transmitter nor the receiver is assumed to have CSI.

The received matrix $\boldsymbol{Y} \in \mathbb{C}^{N\times T}$ is given by

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \gamma^{-1/2}\boldsymbol{W} \tag{1}$$

in one fading block. In (1), the channel matrix $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ and the additive white Gaussian noise (AWGN) matrix $\boldsymbol{W} \in \mathbb{C}^{N\times T}$ are assumed to be independent standard complex Gaussian matrices. The transmitted matrix $\boldsymbol{X} \in \mathbb{C}^{M\times T}$ satisfies the average power constraint $\mathbb{E}[\mathrm{Tr}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}})] \leq T$ [16]. The parameter $\gamma > 0$ in (1) represents the SNR.

B. Capacity and Energy Efficiency

Let $C_{\mathrm{opt}}(\gamma)$ denote the channel capacity in nats for the MIMO channel (1) with no CSI [16]. The EE $C_{\mathrm{opt}}(\gamma)/\gamma$ is defined as the capacity divided by the SNR, which is maximized as the SNR γ tends to zero [12]. Thus, investigating the optimum EE $\sup_{\gamma>0} C_{\mathrm{opt}}(\gamma)/\gamma$ is equivalent to analyzing the capacity $C_{\mathrm{opt}}(\gamma)$ in the low SNR regime $\gamma \to 0$. Thus, we focus on the capacity in the low SNR regime.

It was proved in [14, Theorem 3] that for any $\alpha \in (0, 1]$

$$C_{\mathrm{opt}}(\gamma) = N\gamma - \frac{N(M+N)}{2M}\gamma^{1+\alpha} + o(\gamma^{1+\alpha}) \text{ as } \gamma \to 0, \tag{2}$$

if and only if there is some $\epsilon > 0$ such that T satisfies

$$T = \frac{M^2}{(M+N)^2}\gamma^{-2(\alpha+\epsilon)}. \tag{3}$$

When $\alpha = 1$, the right-hand side (RHS) of (2) is equal to the capacity of the MIMO channel (1) with CSIR up to second order in the low SNR regime [12]. Thus, (3) is regarded as a necessary and sufficient condition for achieving the capacity for the CSIR case up to second order. A signaling scheme with achievable rate $C(\gamma)$ is called second-order optimal [12] when $C(\gamma)$ is equal to the corresponding capacity up to second order ($O(\gamma^{1+\alpha})$ for $\alpha \in (0, 1]$) in the low SNR regime. By definition, the EE of second-order-optimal schemes is equal to the optimum EE $C_{\text{opt}}(\gamma)/\gamma$ up to $O(\gamma^{\alpha})$ as $\gamma \to 0$.

C. Generalized Spatial Modulation

In GSM, the transmitter activates $\tilde{M}$ transmit antennas $\mathcal{M}_t \subset \mathcal{M}_{\text{all}} = \{1, \dots, M\}$ with $|\mathcal{M}_t| = \tilde{M}$ in symbol period t, and send data symbols only from the activated antennas $\mathcal{M}_t$. Thus, the *sparse* vector $\boldsymbol{x}_{\mathcal{M}_t,t} \in \mathbb{C}^M$ transmitted in symbol period t satisfies $x_{m,t} = (\boldsymbol{x}_{\mathcal{M}_t,t})_m = 0$ for all $m \notin \mathcal{M}_t$. We assume that $\{\mathcal{M}_t\}$ are sampled from all possible subsets of $\mathcal{M}_{\text{all}}$ with cardinality $\tilde{M}$ uniformly and randomly for all t, and that $\{\boldsymbol{x}_{\mathcal{M}_t,t}\}$ are independent for all t. The covariance matrix $\boldsymbol{\Sigma}_{\mathcal{M}_t} = \mathbb{E}[\boldsymbol{x}_{\mathcal{M}_t,t}\boldsymbol{x}^{\text{H}}_{\mathcal{M}_t,t}]$ satisfies the constraint $\text{Tr}(\boldsymbol{\Sigma}_{\mathcal{M}_t}) \leq 1$ to fulfill the average power constraint. Furthermore, we assume that the *complex* data symbols $\{x_{m,t}\}$ with finite fourth moments satisfy

$$\mathbb{E}[x_{m_1,t}] = 0, \quad \mathbb{E}[x_{m_1,t}x_{m_2,t}] = 0,$$
$$\mathbb{E}[x_{m_1,t}x_{m_2,t}x_{m_3,t}] = 0, \quad \mathbb{E}\left[x_{m_1,t}x_{m_2,t}x^*_{m_3,t}\right] = 0, \tag{4}$$

for all $m_1, m_2, m_3 \in \mathcal{M}_t$.

The assumption (4) should be regarded as a sufficient condition under which the modulation scheme for the non-zero elements of $\boldsymbol{x}_{\mathcal{M}_t,t}$ is second-order optimal for the equivalent MIMO channel with CSIR,

$$\boldsymbol{y}_t = \sum_{m\in\mathcal{M}_t} \boldsymbol{h}_m x_{m,t} + \gamma^{-1/2}\boldsymbol{w}_t. \tag{5}$$

In (5), $\boldsymbol{y}_t$ and $\boldsymbol{w}_t$ denote the tth columns of $\boldsymbol{Y}$ and $\boldsymbol{W}$ in (1), respectively. The vector $\boldsymbol{h}_m$ represents the mth column of $\boldsymbol{H}$.

For example, the assumption (4) holds for $\boldsymbol{x}_{\mathcal{M}_t,t} = \sqrt{\boldsymbol{\Sigma}_{\mathcal{M}_t}}\boldsymbol{b}_{\mathcal{M}_t,t}$, in which the precoding matrix $\sqrt{\boldsymbol{\Sigma}_{\mathcal{M}_t}}$ is a square root of $\boldsymbol{\Sigma}_{\mathcal{M}_t}$, and in which $\boldsymbol{b}_{\mathcal{M}_t,t} \in \mathbb{C}^M$ has i.i.d. non-zero elements $\{b_{m,t} = (\boldsymbol{b}_{\mathcal{M}_t,t})_m\}$ only for all $m \in \mathcal{M}_t$, with $\mathbb{E}[b_{m,t}] = \mathbb{E}[b^2_{m,t}] = \mathbb{E}[b^3_{m,t}] = \mathbb{E}[b_{m,t}|b_{m,t}|^2] = 0$. Quadrature phase-shift keying (QPSK) is an important example that satisfies the assumptions for the data symbol $b_{m,t}$.

III. Main Results

A. Perfect CSI at the Receiver

We first analyze the coherent achievable rate $C^{\text{CSI}}_{\text{GSM}}(\gamma; \boldsymbol{H})$ of the GSM for a fixed known channel matrix $\boldsymbol{H}$. In Section III-A the channel matrix $\boldsymbol{H}$ is regarded as a deterministic matrix, so that conditioning with respect to $\boldsymbol{H}$ is omitted.

Let $\mathcal{D}_t = \{\boldsymbol{x}_{\mathcal{M}_t,t}, \mathcal{M}_t\}$ denote the transmitted data in symbol period t. The instantaneous achievable rate $C^{\text{CSI}}_{\text{GSM}}(\gamma; \boldsymbol{H})$ is defined as

$$C^{\text{CSI}}_{\text{GSM}}(\gamma; \boldsymbol{H}) = \frac{1}{T}\sum_{t=1}^{T} I(\mathcal{D}_t; \boldsymbol{y}_t). \tag{6}$$

Since the mutual information $I(\mathcal{D}_t; \boldsymbol{y}_t)$ given $\boldsymbol{H}$ is independent of t, for notational convenience, we omit the time index t from all variables in Section III-A. The following theorem is obtained by evaluating (6) in the low SNR regime.

Theorem 1: Suppose that the modulation scheme for the non-zero elements of $\boldsymbol{x}_{\mathcal{M}}$ satisfies the assumption (4). Then, for any $\tilde{M}$ and any realization of the channel matrix $\boldsymbol{H}$

$$C^{\text{CSI}}_{\text{GSM}}(\gamma; \boldsymbol{H}) = \gamma\text{Tr}\left(\boldsymbol{H}\mathbb{E}_{\mathcal{M}}[\boldsymbol{\Sigma}_{\mathcal{M}}]\boldsymbol{H}^{\text{H}}\right) - \frac{\gamma^2}{2}\text{Tr}\left\{\left(\boldsymbol{H}\mathbb{E}_{\mathcal{M}}[\boldsymbol{\Sigma}_{\mathcal{M}}]\boldsymbol{H}^{\text{H}}\right)^2\right\} + o(\gamma^2). \tag{7}$$

Proof: See Appendix A. □

Theorem 1 is a generalization of the result for MISO ($N=1$) channels with SM ($\tilde{M}=1$) and CSIR in [4]. Note that Theorem 1 can be proved with [13, Corollary 2] and $I(\mathcal{D}_t; \boldsymbol{y}_t) = I(\boldsymbol{x}_t; \boldsymbol{y}_t)$, in which the tth column $\boldsymbol{x}_t$ of $\boldsymbol{X}$ follows the mixture distribution $P_{\boldsymbol{x}_t}(\boldsymbol{x}) = \mathbb{E}_{\mathcal{M}_t}[P_{\boldsymbol{x}_{\mathcal{M}_t,t}}(\boldsymbol{x})]$.

Corollary 1: Suppose that the modulation scheme for the non-zero elements of $\boldsymbol{x}_{\mathcal{M}}$ satisfies the assumption (4), and that the non-zero elements $\{x_{m,t}\}$ are independent variables with average power $\mathbb{E}[|x_{m,t}|^2] = \tilde{M}^{-1}$. Then, for any number $\tilde{M}$ of simultaneously activated antennas, the ergodic achievable rate $C^{\text{CSI}}_{\text{GSM}}(\gamma) = \mathbb{E}_{\boldsymbol{H}}[C^{\text{CSI}}_{\text{GSM}}(\gamma; \boldsymbol{H})]$ is given by

$$C^{\text{CSI}}_{\text{GSM}}(\gamma) = N\gamma - \frac{N(M+N)}{2M}\gamma^2 + o(\gamma^2). \tag{8}$$

Proof: It is straightforward to confirm that the $M \times M$ matrix $\mathbb{E}_{\mathcal{M}}[\boldsymbol{\Sigma}_{\mathcal{M}}]$ is equal to the optimum covariance matrix $M^{-1}\boldsymbol{I}_M$ for the original i.i.d. Rayleigh fading MIMO channel (1) with CSIR. See [12] for evaluation of the ergodic achievable rate $C^{\text{CSI}}_{\text{GSM}}(\gamma)$ with $\mathbb{E}_{\mathcal{M}}[\boldsymbol{\Sigma}_{\mathcal{M}}] = M^{-1}\boldsymbol{I}_M$. □

Surprisingly, the achievable rate (8) is independent of $\tilde{M}$ in the low SNR regime. The assumptions in Corollary 1 hold for equal-power QPSK, which is second-order optimal for the original MIMO channel (1) [12, Theorem 14]. Corollary 1 implies that QPSK GSM is second-order optimal for the i.i.d. Rayleigh fading MIMO channel with CSIR, regardless of the number of simultaneously activated transmit antennas.

The optimality of QPSK was shown in terms of error probability in the high SNR regime [17], [18]: QPSK SM achieves the best performance in the high SNR regime if more transmit antennas can be used to increase the transmission rate. Corollary 1 indicates that QPSK is optimal in the low SNR regime, regardless of the number of transmit antennas.

B. No CSI

It is known that it is wasteful to spread the transmit power over time for the no CSI case. An energy-efficient scheme is flash signaling [14], in which data transmission is performed only in a part of all fading blocks to reduce the number of fading blocks to be estimated. Flash signaling concentrates the transmit power on a part of fading blocks. Thus, it should be regarded as power allocation over time.

We propose flash GSM as an energy-efficient scheme for the no CSI case. In flash GSM, transmission is performed only in $\delta(\gamma)$ fraction of all fading blocks. When transmission is active in a fading block, GSM with SNR $\gamma' = \gamma/\delta(\gamma)$ is used to satisfy the average power constraint over all fading blocks.

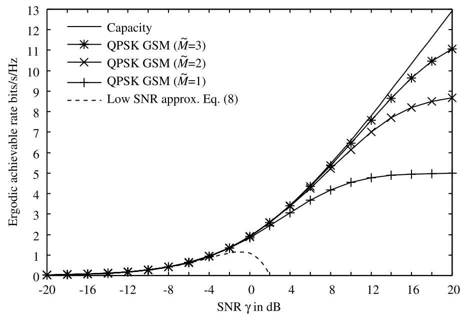

Fig. 1. Ergodic achievable rate versus SNR γ for QPSK GSM over the i.i.d. Rayleigh fading MIMO channel with CSIR. $M = 8$, $N = 2$, and $T = 1$.

In Section III-B averaging over the channel matrix $\boldsymbol{H}$ is considered, whereas $\boldsymbol{H}$ was fixed in Section III-A. Note that conditioning with respect to $\boldsymbol{H}$ is not omitted in Section III-B.

We define the achievable rate $C_{\mathrm{FGSM}}(\gamma)$ of the flash GSM for the MIMO channel (1) with no CSI as

$$C_{\mathrm{FGSM}}(\gamma) = \frac{\delta(\gamma)}{T} I\left(\{\mathcal{D}_t : \text{all } t\}; \boldsymbol{Y}\right). \quad (9)$$

Theorem 2: Suppose that $\boldsymbol{x}_{\mathcal{M}_t,t}$ satisfies all assumptions in Corollary 1. For any $\alpha \in (0, 1]$, the achievable rate $C_{\mathrm{FGSM}}(\gamma)$ of the flash GSM with $\delta(\gamma) = \gamma^{1-\alpha}$ for no CSI is given by

$$C_{\mathrm{FGSM}}(\gamma) = N\gamma - \frac{N(M+N)}{2M}\gamma^{1+\alpha} + o(\gamma^{1+\alpha}), \quad (10)$$

if there is some $\epsilon > 0$ such that T satisfies (3).

Proof: See Appendix B. □

Comparing (2) and (10) implies that the flash GSM is second-order optimal for the MIMO channel (1) with no CSI.

IV. Numerical Results

In order to verify Corollary 1, numerical results are presented for QPSK GSM. Fig. 1 shows the ergodic achievable rates $C_{\mathrm{GSM}}^{\mathrm{CSI}}(\gamma) = \mathbb{E}_{\boldsymbol{H}}[C_{\mathrm{GSM}}^{\mathrm{CSI}}(\gamma; \boldsymbol{H})]$ of the GSM for the CSIR case—estimated by Monte Carlo integration of (6)—as well as its low SNR approximation (8). For comparison, the channel capacity is also plotted.

The low SNR approximation (8) is accurate for $\gamma < -5$ dB. In the same low SNR regime, the achievable rates of the GSM are indistinguishable from the capacity for all $\tilde{M}$, as predicted in Corollary 1. As the maximum transmission rate $R = 2\tilde{M} + \log_2 \binom{M}{\tilde{M}}$ in bits increases, the achievable rates get closer to the capacity in the moderate SNR regime, while they are limited to below R in the high SNR regime.

V. Conclusion

The main results of this letter are twofold: One is that QPSK GSM is second-order optimal for the i.i.d. Rayleigh block-fading MIMO channel with CSIR, regardless of the number of simultaneously activated transmit antennas. The other is that, under a condition for coherence time, flash GSM with QPSK is second-order optimal for the no CSI case. The conclusion of this letter is that single-RF MIMO systems with QPSK are practically the best option in terms of the information-theoretical EE under rich scattering environment.

An interesting future work is to construct precoded GSM for correlated MIMO channels with CSIR or for MIMO channels with full CSI. The low SNR approximation of the achievable rate in Theorem 1 can be used as an objective function for designing energy-efficient precoders in the low SNR regime. The maximization of the objective function over the input covariance matrices is concave, so that the precoder-design problem may be solved efficiently.

Appendix A
Proof of Theorem 1

We shall evaluate $I(\mathcal{D}; \boldsymbol{y}) = h(\boldsymbol{y}) - h(\boldsymbol{y}|\mathcal{D})$ in nats. From (5), $h(\boldsymbol{y}|\mathcal{D})$ given $\boldsymbol{H}$ is equal to the differential entropy of the AWGN vector $\gamma^{-1/2}\boldsymbol{w}$ [15], i.e. $h(\boldsymbol{y}|\mathcal{D}) = N \ln(\pi e \gamma^{-1})$.

The differential entropy $h(\boldsymbol{y})$ given $\boldsymbol{H}$ is defined as

$$h(\boldsymbol{y}) = -\int p(\boldsymbol{y}|\boldsymbol{H}) \ln p(\boldsymbol{y}|\boldsymbol{H}) d\boldsymbol{y}, \quad (11)$$

with

$$p(\boldsymbol{y}|\boldsymbol{H}) = \mathbb{E}_{\mathcal{D}}\left[p(\boldsymbol{y}|\boldsymbol{H}, \mathcal{D})\right], \quad (12)$$

where $p(\boldsymbol{y}|\boldsymbol{H}, \mathcal{D})$ represents the MIMO channel (1) with the GSM in a symbol period,

$$p(\boldsymbol{y}|\boldsymbol{H}, \mathcal{D}) = \frac{1}{(\pi\gamma^{-1})^N} e^{-\gamma\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}}\|^2}. \quad (13)$$

Extracting a $\mathcal{D}$-independent factor from (13) yields

$$p(\boldsymbol{y}|\boldsymbol{H}) = \frac{1}{(\pi\gamma^{-1})^N} e^{-\gamma\|\boldsymbol{y}\|^2} G(\sqrt{\gamma}\boldsymbol{y}; \gamma), \quad (14)$$

with

$$G(\boldsymbol{u}; \gamma) = \mathbb{E}_{\mathcal{D}}\left[\exp\left\{2\sqrt{\gamma}\Re\left[(\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}})^{\mathrm{H}}\boldsymbol{u}\right] - \gamma\|\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}}\|^2\right\}\right]. \quad (15)$$

Substituting (14) into (11) and using the change of variables $\boldsymbol{u} = \sqrt{\gamma}\boldsymbol{y}$, we obtain

$$h(\boldsymbol{y}) = N\ln(\pi e \gamma^{-1}) + \gamma\mathbb{E}_{\mathcal{D}}\left[\|\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}}\|^2\right] + H(\gamma), \quad (16)$$

with

$$H(\gamma) = -\mathbb{E}_{\boldsymbol{u}}\left[G(\boldsymbol{u}; \gamma)\ln G(\boldsymbol{u}; \gamma)\right], \quad (17)$$

where $\boldsymbol{u} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$ is regarded as a standard complex Gaussian vector.

The first term on the RHS of (16) is equal to the differential entropy $h(\boldsymbol{y}|\mathcal{D})$. Furthermore, the second term reduces to the first term on the RHS of (7). Thus, it is sufficient to prove

$$H(\gamma) = -\frac{\gamma^2}{2}\mathrm{Tr}\left\{\left(\boldsymbol{H}\mathbb{E}_{\mathcal{M}}[\boldsymbol{\Sigma}_{\mathcal{M}}]\boldsymbol{H}^{\mathrm{H}}\right)^2\right\} + o(\gamma^2). \quad (18)$$

We follow [4] to expand (17) with respect to $\sqrt{\gamma}$ up to fourth order in the low SNR limit $\gamma \to 0$. Although the derivation is omitted, as a natural generalization of [4], under the assumption (4) we arrive at

$$H(\gamma) = -\gamma\mathbb{E}_{\boldsymbol{u}}[a_1(\boldsymbol{u})] - \gamma^2\mathbb{E}_{\boldsymbol{u}}\left[a_2(\boldsymbol{u}) + \frac{a_1(\boldsymbol{u})^2}{2}\right] + o(\gamma^2), \quad (19)$$

with

$$a_1(\boldsymbol{u}) = 2\mathbb{E}_{\mathcal{D}}\left[\left\{\Re\left[(\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}})^{\mathrm{H}}\boldsymbol{u}\right]\right\}^2\right] - \mathbb{E}_{\mathcal{D}}\left[\|\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}}\|^2\right], \quad (20)$$

$$a_2(\boldsymbol{u}) = \frac{1}{2}\mathbb{E}_{\mathcal{D}}\left[\|\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}}\|^4\right] + \frac{2}{3}\mathbb{E}_{\mathcal{D}}\left[\left\{\Re\left[(\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}})^{\mathrm{H}}\boldsymbol{u}\right]\right\}^4\right] - 2\mathbb{E}_{\mathcal{D}}\left[\left\{\Re\left[(\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}})^{\mathrm{H}}\boldsymbol{u}\right]\right\}^2 \|\boldsymbol{H}\boldsymbol{x}_{\mathcal{M}}\|^2\right]. \quad (21)$$

In order to evaluate (19), we use the two properties

$$\mathbb{E}_{\boldsymbol{u}}\left[\left\{\Re[\boldsymbol{a}^{\mathrm{H}}\boldsymbol{u}]\right\}^2\right] = \frac{1}{2}\|\boldsymbol{a}\|^2, \quad \mathbb{E}_{\boldsymbol{u}}\left[\left\{\Re[\boldsymbol{a}^{\mathrm{H}}\boldsymbol{u}]\right\}^4\right] = \frac{3}{4}\|\boldsymbol{a}\|^4, \quad (22)$$

for any $\boldsymbol{a} \in \mathbb{C}^N$. Exchanging the two expectations $\mathbb{E}_{\mathcal{D}}[\cdots]$ and $\mathbb{E}_{\boldsymbol{u}}[\cdots]$ in (19), we find $\mathbb{E}_{\boldsymbol{u}}[a_i(\boldsymbol{u})] = 0$ for $i = 1, 2$.

Calculating the expectation in (20) with respect to $\boldsymbol{x}_{\mathcal{M}}$ with $\mathbb{E}[x_m x_{m'}] = 0$ for all m and m' yields

$$a_1(\boldsymbol{u}) = \mathrm{Tr}\left\{\boldsymbol{H}\mathbb{E}_{\mathcal{M}}[\boldsymbol{\Sigma}_{\mathcal{M}}]\boldsymbol{H}^{\mathrm{H}}(\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}} - \boldsymbol{I}_N)\right\}. \quad (23)$$

Evaluating[1] the average of $a_1(\boldsymbol{u})^2$ over $\boldsymbol{u} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$, we obtain $\mathbb{E}_{\boldsymbol{u}}[a_1(\boldsymbol{u})^2] = \mathrm{Tr}\{(\boldsymbol{H}\mathbb{E}_{\mathcal{M}}[\boldsymbol{\Sigma}_{\mathcal{M}}]\boldsymbol{H}^{\mathrm{H}})^2\}$. From these observations, we find that (19) reduces to (18) to be proved.

APPENDIX B
PROOF OF THEOREM 2

The proof is an application of [14]. Using the chain rule [15] for mutual information, from (9) we obtain

$$\begin{aligned} C_{\mathrm{FGSM}}(\gamma) &= (\delta(\gamma)/T)\left\{I\left(\{\mathcal{D}_t\}, \boldsymbol{H}; \boldsymbol{Y}\right) - I\left(\boldsymbol{H}; \boldsymbol{Y}|\{\mathcal{D}_t\}\right)\right\} \\ &= \delta(\gamma)\left\{C_{\mathrm{GSM}}^{\mathrm{CSI}}(\gamma') + T^{-1}\left[I(\boldsymbol{H}; \boldsymbol{Y}) - I\left(\boldsymbol{H}; \boldsymbol{Y}|\{\mathcal{D}_t\}\right)\right]\right\} \\ &\geq \delta(\gamma)C_{\mathrm{GSM}}(\gamma'), \end{aligned} \quad (25)$$

with

$$C_{\mathrm{GSM}}(\gamma') = C_{\mathrm{GSM}}^{\mathrm{CSI}}(\gamma') - \frac{1}{T}I\left(\boldsymbol{H}; \boldsymbol{Y}|\{\mathcal{D}_t\}\right), \quad (26)$$

where $C_{\mathrm{GSM}}^{\mathrm{CSI}}(\gamma')$ denotes the ergodic coherent achievable rate $T^{-1}I(\{\mathcal{D}_t\}; \boldsymbol{Y}|\boldsymbol{H})$ for SNR γ'. In (26), the second term is interpreted as the penalty due to channel uncertainty.

Assume $\alpha = 1$ (i.e. $\delta(\gamma) = 1$ and $\gamma' = \gamma$). Since $\boldsymbol{H}$ is a standard complex Gaussian matrix, we find that the second term on the RHS of (26) reduces to

$$\frac{1}{T}I\left(\boldsymbol{H}; \boldsymbol{Y}|\{\mathcal{D}_t\}\right) = \frac{N}{T}\mathbb{E}\left[\ln\det\left(\boldsymbol{I}_M + \gamma\sum_{t=1}^{T}\boldsymbol{\Xi}_{\mathcal{M}_t,t}\right)\right], \quad (27)$$

with $\boldsymbol{\Xi}_{\mathcal{M}_t,t} = \boldsymbol{x}_{\mathcal{M}_t,t}\boldsymbol{x}_{\mathcal{M}_t,t}^{\mathrm{H}}$. Using Jensen's inequality yields

$$\frac{1}{T}I\left(\boldsymbol{H}; \boldsymbol{Y}|\{\mathcal{D}_t\}\right) \leq \frac{MN}{T}\ln\left(1 + \frac{T\gamma}{M}\right), \quad (28)$$

where we have used $\mathbb{E}_{\mathcal{D}_t}[\boldsymbol{x}_{\mathcal{M}_t,t}\boldsymbol{x}_{\mathcal{M}_t,t}^{\mathrm{H}}] = M^{-1}\boldsymbol{I}_M$.

[1]For i.i.d. CSCG random variables $u_i \sim \mathcal{CN}(0, 1)$ with unit variance, use

$$\mathbb{E}\left[u_i u_j u_k^* u_l^*\right] = \begin{cases} 2 & \text{for } i = j = k = l, \\ 1 & \text{for } (i, j) = (k, l) \text{ or } (i, j) = (l, k), \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

The upper bound (28) is $o(\gamma^2)$ if (3) holds for some $\epsilon > 0$ [14, Eq. (9)]. From (25) and (26) we obtain the lower bound $C_{\mathrm{FGSM}}(\gamma) \geq C_{\mathrm{GSM}}^{\mathrm{CSI}}(\gamma) + o(\gamma^2)$. Combining the lower bound and the trivial upper bound $C_{\mathrm{FGSM}}(\gamma) \leq C_{\mathrm{opt}}(\gamma)$ given by (2), from Corollary 1 we arrive at (10) for $\alpha = 1$.

Assume $\alpha \in (0, 1)$. We have already proved that

$$C_{\mathrm{GSM}}(\gamma') = N\gamma' - \frac{N(M+N)}{2M}\gamma'^2 + o(\gamma'^2), \quad (29)$$

if $T = M^2(M+N)^{-2}\gamma'^{-2(1+\epsilon')}$ for some $\epsilon > 0$. From $\delta(\gamma) = \gamma^{1-\alpha}$ and $\gamma' = \gamma/\delta(\gamma)$, this is equivalent to

$$\delta(\gamma)C_{\mathrm{GSM}}(\gamma') = N\gamma - \frac{N(M+N)}{2M}\gamma^{1+\alpha} + o(\gamma^{1+\alpha}), \quad (30)$$

if $T = M^2(M+N)^{-2}\gamma^{-2\alpha(1+\epsilon')}$. Thus, from (25) we obtain Theorem 2 by letting $\epsilon = \alpha\epsilon'$.

REFERENCES

[1] R. Y. Mesleh, H. Haas, S. Sinanović, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.
[2] A. Stavridis, S. Sinanović, M. Di Renzo, and H. Haas, "Energy evaluation of spatial modulation at a multi-antenna base station," in *Proc. IEEE 78th Veh. Technol. Conf.*, Las Vegas, NV, USA, Sep. 2013, pp. 1–5.
[3] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, oppotunities, and implementation," *Proc. IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.
[4] K. Takeuchi, "Spatial modulation achieves information-theoretically optimal energy efficiency," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1133–1136, Jul. 2015.
[5] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Generalized space shift keying modulation for MIMO channels," in *Proc. IEEE 19th Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Cannes, France, Sep. 2008, pp. 1–5.
[6] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalised spatial modulation," in *Proc. 44th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2010, pp. 1498–1502.
[7] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3219–3230, Nov. 2010.
[8] M. Di Renzo, D. D. Leonardis, F. Graziosi, and H. Haas, "Space Shift Keying (SSK-) MIMO with practical channel estimates," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 998–1012, Apr. 2012.
[9] P. Zhang, S. Chen, and L. Hanzo, "Reduced-complexity near-capacity joint channel estimation and three-stage turbo detection for coherent space-time shift keying," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1902–1913, May 2013.
[10] 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.
[11] 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.
[12] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.
[13] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.
[14] S. Ray, M. Médard, and L. Zheng, "On noncoherent MIMO channels in the wideband regime: Capacity and reliability," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 1983–2009, Jun. 2007.
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[16] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
[17] 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.
[18] M. Maleki, H. R. Bahrami, A. Alizadeh, and N. H. Tran, "On the performance of spatial modulation: Optimal constellation breakdown," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 144–157, Jan. 2014.