

Generalized Spatial Modulation Aided MmWave MIMO with Sub-Connected Hybrid Precoding Scheme

Longzhuang He, Jintao Wang, and Jian Song

Department of Electronic Engineering, Tsinghua University, Beijing, China

Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing, China

Email: helongzhuang@126.com

Abstract—Due to the high cost and low energy efficiency of the dedicated radio frequency (RF) chains, the number of RF chains in a millimeter wave (mmWave) multiple-input multiple-output (MIMO) system is usually limited from a practical point of view. In this case, the maximum number of independent data streams is also restricted by the number of RF chains, which consequently leads to limiting the potentially attainable spatial multiplexing gain. In order to address this issue, in this paper, a novel generalized spatial modulation (GenSM) aided mmWave MIMO system is proposed, which enables the transmission of an extra data stream via the index of the active antennas group and requires no extra RF chain. Moreover, a two-step algorithm is also proposed to optimize the hybrid precoder design with respect to spectral efficiency (SE) maximization. Finally, numerical simulation results demonstrate the superior SE performance achieved by the proposed scheme.

Index Terms—Generalized spatial modulation; multiple-input multiple-output (MIMO); millimeter wave (mmWave); hybrid precoding; convex optimization; ℓ_∞ norm; spectral efficiency.

I. INTRODUCTION

The idea of millimeter wave (mmWave) communication has recently attracted substantial research interest due to its potential to significantly boost the throughput for future communication networks [1]. By incorporating with multiple-input multiple-output (MIMO) techniques, mmWave MIMO is capable of striking a substantial spectral efficiency (SE) gain via precoding at the transmitter [2].

Due to the high cost and low energy efficiency of the radio frequency (RF) chains, the number of RF chains in mmWave MIMOs is usually limited from a practical point of view, which facilitates the application of the well-known *hybrid precoding* [3][4]. While hybrid precoding achieves a near-optimal performance with a limited number of RF chains [3], the maximum number of independent data streams is still restricted by the number of RF chains, which therefore limits the potential spatial multiplexing (SMX) gain. In order to address this issue, the concepts of spatial modulation (SM) and generalized SM (GenSM) have been recently incorporated with mmWave MIMOs in [5]-[7]. In SM/GenSM systems [8]-[10], a subset of antennas are randomly activated to transmit classic amplitude-phase modulation (APM) symbols, and the information is thus simultaneously conveyed via the APM symbols and the indices of the active antennas.

However, the previous research on the combination of SM/GenSM and mmWave MIMO, i.e. [5] and [6], all failed to fully exploit the transmitter's knowledge of the channel state information (CSI) for an optimized precoder design, which leads to severe performance penalty when compared to the optimal MIMO waterfilling capacity. To the best of the authors' knowledge, the design of hybrid precoding in a GenSM-aided mmWave MIMO scenario has not been explored.

In order to address this issue, the major contributions of this paper are summarized as follows.

- 1) A novel GenSM-aided mmWave MIMO scheme with a sub-connected hybrid precoding structure is proposed, where an extra data stream is conveyed via the index of the active antennas group without requiring any extra RF chains.
- 2) A closed-form SE lower bound is proposed to quantify the achievable SE of the proposed scheme, which is shown to provide a favorable approximation accuracy with a constant shift.
- 3) Using the proposed bound as the optimization target, a two-step algorithm is proposed to iteratively optimize the hybrid precoder in terms of SE maximization. As our proposed scheme constitutes a more generalized hybrid precoding paradigm, substantial SE gain over the conventional precoding schemes is thus achieved, as witnessed by our numerical simulations.

The organization of this paper is summarized as follows. Section II introduces the system model of our proposed GenSM-aided mmWave MIMO. Theoretical SE analysis is provided in Section III. Section IV introduces our proposed two-step optimization algorithm. The simulation results are provided in Section V, while Section VI concludes this paper.

Notations: The lowercase and uppercase boldface letters denote column vectors and matrices respectively. The operators $(\cdot)^T$ and $(\cdot)^H$ denote the transposition and conjugate transposition, respectively. The determinant of a matrix \mathbf{M} is denoted by $|\mathbf{M}|$, while \otimes represents the Kronecker product.

II. SYSTEM MODEL

We depict our proposed GenSM-aided mmWave MIMO scheme in Fig.1, where an mmWave MIMO with N_T transmit antennas (TAs) and N_R receive antennas (RAs) are considered.

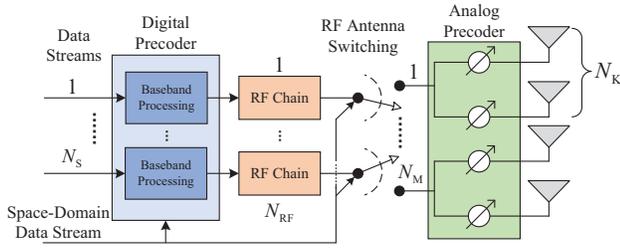


Fig. 1. Block diagrams of the proposed GenSM-aided mmWave MIMO scheme.

As depicted in the figure, the digital precoder, which consists of N_s baseband processing units, essentially plays the role of conducting *power allocation* to the N_s input data streams. Note that the power allocation vector simultaneously depends on the input space-domain data stream as well as the instantaneous CSI. This space-information-guided digital precoding is practically achievable due to the high-speed and low-latency advantages of based digital processing.

Moreover, the N_T TAs are grouped into N_M antenna groups (AGs), each of which consists of N_K TAs with $N_T = N_M N_K$. It is required that $N_M \geq N_{RF}$, and the space-domain information randomly assigns the RF-domain symbols to N_{RF} out of the N_M AGs. Finally, similar to [4], N_K phase shifters (PSs) are equipped in each AG to perform analog precoding. Note that the conventional sub-connected hybrid precoding scheme [4] can be conceived as a special case of the proposed framework, when $N_M = N_{RF}$.

The input space-domain information plays the role of specifying an AG activation pattern, which corresponds to a combination of active AGs (AGC). Therefore the total number of legitimate AGCs are given by $M = 2^{\lfloor \log_2 \binom{N_M}{N_{RF}} \rfloor}$ [9], where $\lfloor \cdot \rfloor$ represents the floor operation, and $\binom{\cdot}{\cdot}$ is the binomial coefficient. By denoting the AGs' indices activated by the m -th AGC ($1 \leq m \leq M$) as $\mathbf{u}_m \triangleq [u_{m1}, u_{m2}, \dots, u_{mN_{RF}}]^T$ with $1 \leq u_{mn} \leq N_M$, the m -th *AG-selection matrix* $\mathbf{C}_m \in \mathbb{R}_{N_T \times N_{RF}}$ can therefore be given as follows ($1 \leq m \leq M$):

$$\mathbf{C}_m \triangleq [\mathbf{e}_{u_{m1}}, \mathbf{e}_{u_{m2}}, \dots, \mathbf{e}_{u_{mN_{RF}}}] \otimes \mathbf{1}_{N_K}, \quad (1)$$

where $\mathbf{1}_{N_K} \in \mathbb{R}_{N_K \times 1}$ is an N_K -dimensional all-one vector and \mathbf{e}_m represents the m -th column of \mathbf{I}_{N_M} .

The digital precoder is composed of M diagonal real-valued matrices, i.e. $\mathcal{D} \triangleq \{\mathbf{D}_1, \dots, \mathbf{D}_M\}$, where $\mathbf{D}_m = \text{diag}(d_{m1}, d_{m2}, \dots, d_{mN_s}) \in \mathbb{R}_{N_s \times N_s}$ is applied, when the m -th AGC is selected. Finally, similar to [4], the analog precoder $\mathbf{A} \in \mathbb{C}_{N_T \times N_T}$ is given by:

$$\mathbf{A} \triangleq \text{diag} \left(\frac{1}{\sqrt{N_K}} e^{j\theta_1}, \frac{1}{\sqrt{N_K}} e^{j\theta_2}, \dots, \frac{1}{\sqrt{N_K}} e^{j\theta_{N_T}} \right), \quad (2)$$

with $\theta_n \in [0, 2\pi)$ denoting the rotation angle of the n -th TA.

Therefore the received signal vector $\mathbf{y} \in \mathbb{C}_{N_R \times 1}$ can be formulated as follows, when the m -th AGC is selected:

$$\mathbf{y} = \sqrt{\rho} \mathbf{H} \mathbf{A} \mathbf{C}_m \mathbf{D}_m \mathbf{x} + \mathbf{n}, \quad (3)$$

where a narrowband MIMO channel matrix $\mathbf{H} \in \mathbb{C}_{N_R \times N_T}$ is considered, and $\mathbf{x} \sim \mathcal{CN}(0, \frac{1}{N_s} \mathbf{I}_{N_s})$ represents the Gaussian-distributed input signal vector. The average transmit power is given by $\rho > 0$, while $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_N^2 \mathbf{I}_{N_R})$ represents the additive white Gaussian noise (AWGN). Besides, we also assume that the digital precoder satisfies the following power constraint:

$$\sum_{m=1}^M \text{Tr}(\mathbf{D}_m \mathbf{D}_m^H) \leq M N_s, \quad (4)$$

which leads to a signal-to-noise ratio (SNR) of ρ/σ_N^2 at the receiver.

In this paper we adopt the classic clustered Saleh-Valenzuela geometric mmWave channel model [3], which is given as:

$$\mathbf{H} = \gamma \sum_{p=1}^{N_{cl}} \sum_{q=1}^{N_{ray}} \alpha_{pq} \Lambda_t(\phi_{pq}^t, \theta_{pq}^t) \Lambda_r(\phi_{pq}^r, \theta_{pq}^r) \dots \mathbf{b}_t(\phi_{pq}^t, \theta_{pq}^t) \mathbf{b}_r(\phi_{pq}^r, \theta_{pq}^r), \quad (5)$$

where $\gamma > 0$ is the normalizing factor ensuring $E\{\|\mathbf{H}\|_F^2\} = N_R N_T$. For the purpose of brevity, we refer the interested readers to [3] for more details on the specific channel parameters. In this paper, the parameters in (5) will be specified in accordance to [3], unless mentioned otherwise.

III. ACHIEVABLE SPECTRAL EFFICIENCY ANALYSIS

A. Mutual Information Analysis

The achievable SE of the proposed system is quantified by the mutual information (MI) between \mathbf{y} , \mathbf{x} and m , i.e.

$$R(\mathbf{H}, \mathcal{D}, \mathbf{A}) = I(\mathbf{y}; \mathbf{x}, m). \quad (6)$$

However, the MI term in (6) cannot be expressed in a *closed form* due to the discrete channel input $m \in \{1, 2, \dots, M\}$. In order to provide a computationally efficient approach to quantify (6), we propose Theorem 1 to provide a closed-form expression $R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A})$ for lower-bounding $R(\mathbf{H}, \mathcal{D}, \mathbf{A})$:

Theorem 1: The MI term in (6) can be lower-bounded by the following expression:

$$R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A}) = \log_2 \frac{M}{(\sigma_N^2)^{N_R}} \dots - \frac{1}{M} \sum_{n=1}^M \log_2 \sum_{t=1}^M |\Sigma_n + \Sigma_t|^{-1}, \quad (7)$$

where $\Sigma_n \in \mathbb{C}_{N_R \times N_R}$ is given as ($n = 1, 2, \dots, M$):

$$\Sigma_n \triangleq \sigma_N^2 \mathbf{I}_{N_R} + \frac{\rho}{N_s} \mathbf{H} \mathbf{A} \mathbf{C}_n \mathbf{D}_n \mathbf{D}_n^H \mathbf{C}_n^H \mathbf{A}^H \mathbf{H}^H. \quad (8)$$

Proof: See the Appendix A of [11]. ■

B. Bound Tightness

We now seek to demonstrate the tightness of the proposed lower bound in Theorem 1. More specifically, we found that a constant gap exists between the lower bound and the true SE value, when an asymptotically high (low) SNR value is imposed, which is verified by the following proposition.

TABLE I
SIMULATION PARAMETERS

Symbols	Specifications	Typical Values
N_T	Number of TAs	8
N_R	Number of RAs	8
N_K	Number of TAs in each antenna group	2
N_M	Number of antenna groups	4
N_{RF}	Number of RF chains	2
N_S	Number of APM-domain data streams	N_{RF}
λ	Carrier's wavelength	5 mm

Proposition 1: With an asymptotically high or low SNR value, a constant gap of $N_R(1 - \log_2 e)$ exists between the closed-form lower bound $R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A})$ and the true SE expression $R(\mathbf{H}, \mathcal{D}, \mathbf{A})$.

Proof: See the Appendix B of [11]. ■

By adding this constant gap, $R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A})$ is expected to provide a more accurate approximation to $R(\mathbf{H}, \mathcal{D}, \mathbf{A})$. In order to provide numerical demonstrations, we summarize the typical simulation parameters in Table I. Besides, as the hybrid precoder design has not yet been discussed, we tentatively assume that $\mathcal{D} = \{\mathbf{I}_{N_S}, \dots, \mathbf{I}_{N_S}\}$ and $\mathbf{A} = \mathbf{I}_{N_T}/\sqrt{N_K}$.

We depict the true SE expression $R(\mathbf{H}, \mathcal{D}, \mathbf{A})$ and the SE lower bound $R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A})$ in Fig.2. As it can be seen from the figure, aided with the constant shift, the proposed lower bound R_{LB} actually provides a favorable approximation accuracy (the symbols and solid lines are observed to be very close to each other). Since this constant shift is independent from the specific hybrid precoder, applying it thus imposes no impact on the hybrid precoder design in terms of SE maximization, which motivates us to employ R_{LB} as the optimization target for the precoder design.

IV. PROPOSED ALGORITHM FOR PRECODER DESIGN

Since R_{LB} is capable of providing an accurate approximation to the true SE value, we therefore seek to solve the following optimization (P1) to design the hybrid precoder:

$$(P1): \max_{\mathcal{D}, \mathbf{A}} R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A})$$

$$\text{s.t. } \mathcal{D} = \{\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_M\},$$

$$\sum_{m=1}^M \text{Tr}(\mathbf{D}_m \mathbf{D}_m^H) \leq MN_S, \text{ and } \mathbf{A} \in \mathcal{A}, \quad (9)$$

where \mathcal{A} is the feasible set of \mathbf{A} satisfying the definition in (2). A simultaneous optimization of the precoders \mathcal{D} and \mathbf{A} could be computationally infeasible. In order to facilitate a computationally efficient solution, similar to [12], we seek to decompose (P1) and iteratively solve the following two sub-

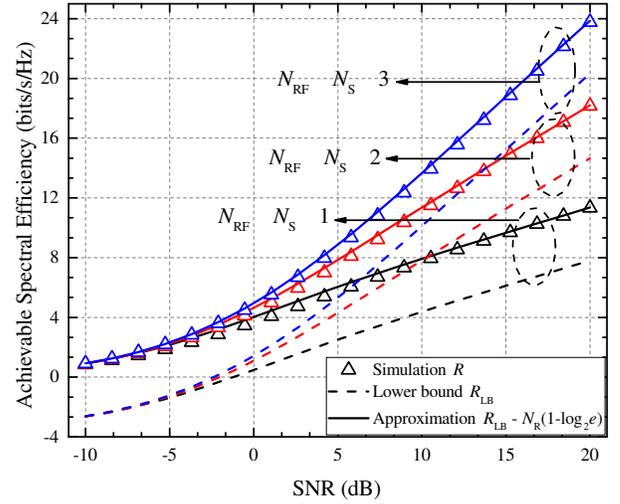


Fig. 2. Average achievable SE and SE bounds with various (N_S, N_{RF}) . The simulation value R and the closed-form lower bound R_{LB} are given according to (6) and (7), respectively. The approximation is given by $[R_{LB} - N_R(1 - \log_2 e)]$.

problems:

$$(P2): \max_{\mathcal{D}} R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A})$$

$$\text{s.t. } \mathcal{D} = \{\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_M\}$$

$$\sum_{m=1}^M \text{Tr}(\mathbf{D}_m \mathbf{D}_m^H) \leq M \cdot N_S, \quad (10)$$

$$(P3): \max_{\mathbf{A}} R_{LB}(\mathbf{H}, \mathcal{D}, \mathbf{A}) \text{ s.t. } \mathbf{A} \in \mathcal{A},$$

in which (P2) optimizes the digital precoder \mathcal{D} based on a given \mathbf{A} , and (P3) optimizes the analog precoder \mathbf{A} upon an given \mathcal{D} .

Before providing our solution to (P2), it is worth pointing out that the cost function R_{LB} is non-concave over $\mathbf{D}_m = \text{diag}(d_{m1}, d_{m2}, \dots, d_{mN_S})$, which can be easily verified via a counterexample with $N_R = N_T = N_K = N_M = N_{RF} = M = 1$. However, the following proposition shows that R_{LB} is actually a concave function of $\boldsymbol{\lambda} = [\lambda_1^T, \dots, \lambda_M^T]^T$ with $\lambda_m \triangleq [d_{m1}^2, d_{m2}^2, \dots, d_{mN_S}^2]^T$.

Proposition 2: The closed-form lower bound R_{LB} is a concave function of the power allocation vector $\boldsymbol{\lambda}$.

Proof: See the Appendix C of [11]. ■

Aided with Proposition 2 and the incorporation of a logarithmic barrier function [13], we therefore seek to solve the following problem:

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^{MN_S \times 1}} f_B(\boldsymbol{\lambda}) = R_{LB}(\boldsymbol{\lambda}) + \sum_{i=1}^{MN_S} \frac{\ln(\lambda_i)}{t_B} \text{ s.t. } \mathbf{1}^T \boldsymbol{\lambda} = MN_S, \quad (12)$$

where λ_i denotes the i -th element of $\boldsymbol{\lambda}$, and t_B is used to scale the penalty of the logarithmic barrier function. The gradient of the cost function in (12) over $\boldsymbol{\lambda}_m$ is thus given as follows:

$$\nabla_{\boldsymbol{\lambda}_m} f_B(\boldsymbol{\lambda}) = \nabla_{\boldsymbol{\lambda}_m} R_{LB}(\boldsymbol{\lambda}) + \frac{1}{t_B} [\lambda_{m1}^{-1}, \dots, \lambda_{mN_S}^{-1}]^T, \quad (13)$$

$$\begin{aligned} \nabla_{\lambda_m} R_{\text{LB}}(\boldsymbol{\lambda}) &= \frac{\rho \log_2 e}{MN_S} \sum_{n=1}^M \frac{|\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_m|^{-1} \text{diag} \left[\mathbf{C}_m^H \mathbf{A}^H \mathbf{H}^H (\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_m)^{-1} \mathbf{H} \mathbf{A} \mathbf{C}_m \right]}{\sum_{t=1}^M |\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_t|^{-1}} + \dots \\ &\frac{\rho \log_2 e}{MN_S} \frac{\sum_{t=1}^M |\boldsymbol{\Sigma}_m + \boldsymbol{\Sigma}_t|^{-1} \text{diag} \left[\mathbf{C}_m^H \mathbf{A}^H \mathbf{H}^H (\boldsymbol{\Sigma}_m + \boldsymbol{\Sigma}_t)^{-1} \mathbf{H} \mathbf{A} \mathbf{C}_m \right]}{\sum_{t=1}^M |\boldsymbol{\Sigma}_m + \boldsymbol{\Sigma}_t|^{-1}}. \end{aligned} \quad (11)$$

where the gradient vector $\nabla_{\lambda_m} R_{\text{LB}}(\boldsymbol{\lambda})$ is given in (11). Moreover, in order to preserve the linear constraint $\mathbf{1}^T \boldsymbol{\lambda} = MN_S$, we thus employ the following ascent direction for λ_m , so that $\mathbf{1}^T \Delta \boldsymbol{\lambda}_m = 0$:

$$\Delta \boldsymbol{\lambda}_m = \left(\mathbf{I}_{MN_S} - \frac{\mathbf{1} \cdot \mathbf{1}^T}{MN_S} \right) \nabla_{\lambda_m} f_B(\boldsymbol{\lambda}). \quad (14)$$

Finally, the ascent direction for $\boldsymbol{\lambda}$ is given as:

$$\Delta \boldsymbol{\lambda} = [\Delta \boldsymbol{\lambda}_1^T, \dots, \Delta \boldsymbol{\lambda}_M^T]^T. \quad (15)$$

Using (15) as the search direction, the global optimal¹ power allocation vector $\boldsymbol{\lambda}^*$ can thus be obtained via a gradient ascent method. The optimal digital precoder $\mathcal{D}^* = \{\mathbf{D}_1^*, \dots, \mathbf{D}_M^*\}$ is thus given by:

$$\mathbf{D}_m^* = \text{diag} \left(\sqrt{\lambda_{m1}^*}, \sqrt{\lambda_{m2}^*}, \dots, \sqrt{\lambda_{mN_S}^*} \right), \quad 1 \leq m \leq M, \quad (16)$$

where $\boldsymbol{\lambda}^* = [(\lambda_1^*)^T, \dots, (\lambda_M^*)^T]^T$.

As for (P3), it is found that (P3) is also a non-concave optimization problem, due to the non-convex constraint ($\mathbf{A} \in \mathcal{A}$) and the non-concavity of R_{LB} over \mathbf{A} . In order to ensure local convergence, we propose to relax (P3) into the following optimization (P3-R) with a convex ℓ_∞ constraint:

$$\text{(P3-R): } \max_{\mathbf{a} \in \mathbb{C}^{N_T \times 1}} R_{\text{LB}}(\mathbf{a}) \text{ s.t. } \|\mathbf{a}\|_\infty \leq 1/\sqrt{N_K}, \quad (18)$$

where $\mathbf{a} = \text{diag}(\mathbf{A}) \in \mathbb{C}^{N_T \times 1}$ denotes the diagonal elements of \mathbf{A} . The ℓ_∞ norm is defined as $\|\mathbf{a}\|_\infty = \max_{n=1, \dots, N_T} |a_n|$.

In order to deal with the non-differentiable ℓ_∞ constraint in (P3-R), similar to [10], we use ℓ_p norm to approximate the ℓ_∞ norm with a large p , which leads to formulating the following optimization with a logarithmic barrier:

$$\max_{\mathbf{a} \in \mathbb{C}^{N_T \times 1}} g_B(\mathbf{a}, p) = R_{\text{LB}}(\mathbf{a}) + \frac{1}{t_B} \ln \left(\frac{1}{\sqrt{N_K}} - \|\mathbf{a}\|_p \right). \quad (19)$$

To solve (19) via a gradient ascent method, we use the gradient as the ascent direction, i.e. $\Delta \mathbf{a} = \nabla_{\mathbf{a}} g_B(\mathbf{a}, p)$. The gradient $\nabla_{\mathbf{a}} g_B(\mathbf{a}, p)$ is given as follows:

$$\nabla_{\mathbf{a}} g_B(\mathbf{a}, p) = \nabla_{\mathbf{a}} R_{\text{LB}}(\mathbf{a}) - \frac{\|\mathbf{a}\|_p^{1-p} \mathbf{p}_a}{2 \cdot t_B \left(N_K^{-1/2} - \|\mathbf{a}\|_p \right)}, \quad (20)$$

where $\nabla_{\mathbf{a}} R_{\text{LB}}(\mathbf{a})$ is given by (17) with $\boldsymbol{\Lambda}_n = \text{diag}(\boldsymbol{\lambda}_n)$, and

¹The global optimality is guaranteed due to the concavity of R_{LB} over $\boldsymbol{\lambda}$, as verified by Proposition 2

$\mathbf{p}_a \in \mathbb{C}^{N_T \times 1}$ is given as:

$$\mathbf{p}_a = \left[a_1 \cdot |a_1|^{p-2}, a_2 \cdot |a_2|^{p-2}, \dots, a_{N_T} \cdot |a_{N_T}|^{p-2} \right]^T. \quad (21)$$

Finally, we present our proposed two-step algorithm as follows, where the digital precoder \mathcal{D} and the analog precoder \mathbf{A} are optimized iteratively.

1. Initialization: Given initial solutions $\mathbf{a}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$. Set iteration index to $i = 0$.

2. Digital Precoder Optimization: Based on $\mathbf{a}^{(i)}$, optimize the digital precoder by solving (12) with a gradient ascent method, which yields $\boldsymbol{\lambda}^{(i+1)}$.

3. Analog Precoder Optimization: Based on $\boldsymbol{\lambda}^{(i+1)}$, optimize the analog precoder by solving (19) with a gradient ascent method, which yields $\mathbf{a}^{(i+1)}$.

4. Halt or Update: Let $i \leftarrow i + 1$. Go to Step 2 until convergence.

We refer the interested readers to [11] for more details on the proposed algorithm. Note that, due to the global convergence of solving (12) and the local convergence of solving (19), the proposed algorithm thus ensures local convergence to a feasible hybrid precoder.

Finally, we focus on the optimized selection of the parameters (N_K, N_M) , when the product $N_T = N_K N_M$ is invariant. Qualitatively, increasing N_K leads to reducing N_M , which therefore reduces the SE gain provided by GenSM, as $M = 2^{\lfloor \log_2 \left(\frac{N_M}{N_{\text{RF}}} \right) \rfloor}$. However, increasing N_K also leads to enlarging the antenna group and thus more array gain can be harnessed by the analog precoder. Hence (N_K, N_M) is the key to achieve a scalable tradeoff between the multiplexing gain and the array gain. We thus seek to optimize the corresponding parameters for maximizing the *average SE*, i.e.

$$\begin{aligned} &(N_K^*, N_M^*) \\ &= \arg \max_{(N_K, N_M)} E_{\mathbf{H}} \{ R_{\text{LB}}[\mathbf{H}, \mathcal{D}^*(\mathbf{H}, N_K, N_M), \dots, \mathbf{A}^*(\mathbf{H}, N_K, N_M)] \}, \text{ s.t. } N_K N_M = N_T, \end{aligned} \quad (22)$$

where $\mathcal{D}^*(\mathbf{H}, N_K, N_M)$ and $\mathbf{A}^*(\mathbf{H}, N_K, N_M)$ denote the hybrid precoder designed by the proposed algorithm based on given \mathbf{H} , N_K and N_M . Here we use R_{LB} instead of R as the cost function in (22) so that a lower computational complexity can be achieved.

Remark: It is worth noting that the conventional hybrid precoding schemes can be conceived as special cases of the proposed framework, when $(N_K, N_M) = (N_T/N_{\text{RF}}, N_{\text{RF}})$. Therefore the proposed scheme has the potential to outperform

$$\nabla_{\mathbf{a}} R_{\text{LB}}(\mathbf{a}) = \frac{\rho \log_2 e}{MN_S} \frac{\sum_{n=1}^M \sum_{t=1}^M |\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_t|^{-1} \text{diag} \left[\mathbf{H}^H (\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_t)^{-1} \mathbf{H} \mathbf{A} (\mathbf{C}_n \boldsymbol{\Lambda}_n \mathbf{C}_n^H + \mathbf{C}_t \boldsymbol{\Lambda}_t \mathbf{C}_t^H) \right]}{\sum_{t'=1}^M |\boldsymbol{\Sigma}_n + \boldsymbol{\Sigma}_{t'}|^{-1}}, \quad (17)$$

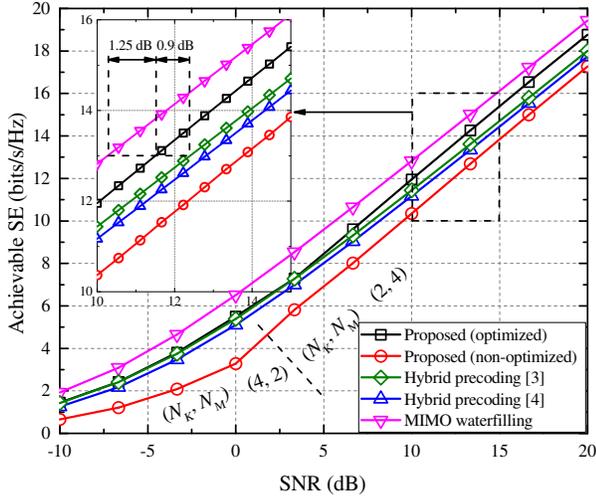


Fig. 3. Achievable SE performance yielded by various schemes with $N_T = 8$, $N_R = 6$ and $N_{\text{RF}} = 2$. The parameters (N_K, N_M) in the proposed GenSM-aided mmWave MIMO scheme is designed by solving (22).

the conventional schemes, given that (N_K, N_M) are optimized according to (22).

V. SIMULATION RESULTS

In this section we present the simulated SE performance yielded by various schemes. Note that the achievable SE performance of the proposed scheme is given by the *true SE expression* $R(\mathbf{H}, \mathcal{D}, \mathbf{A})$ averaged over 1,000 random channel realizations. The simulation parameters are specified in accordance to Table I, unless mentioned otherwise.

In Fig.3, the SE performance yielded by various schemes in conjunction with $(N_T, N_R, N_{\text{RF}}) = (8, 6, 2)$ are presented. It can be observed from the figure that $(N_K, N_M) = (4, 2)$ and $(N_K, N_M) = (2, 4)$ are respectively selected for the proposed scheme (with optimized precoders), when $\text{SNR} < 2.5$ dB and $\text{SNR} > 2.5$ dB. It is also observed that a significant SE gain is achieved by the proposed scheme with an optimized hybrid precoder compared to the non-optimized counterpart, which substantiates the efficacy of the proposed algorithm. By comparing the proposed scheme with state-of-the-art mmWave schemes, it is seen that our scheme maintains a superior SE performance over [4] for the entire SNR range considered, and outperforms the scheme of [3] when $\text{SNR} > 2.5$ dB. Finally, with a target throughput of 13 bits/s/Hz, our scheme outperforms [3] by about 0.9 dB, and is outperformed by the MIMO waterfilling capacity with a 1.25 dB performance gap.

In Fig.4, we increase the number of RF chains to $N_{\text{RF}} = 3$ in a 15×10 mmWave MIMO system, and observe that $(N_K, N_M) = (5, 3)$ and $(N_K, N_M) = (3, 5)$ are respectively selected when $\text{SNR} < -7.5$ dB and $\text{SNR} > -7.5$ dB. Similar

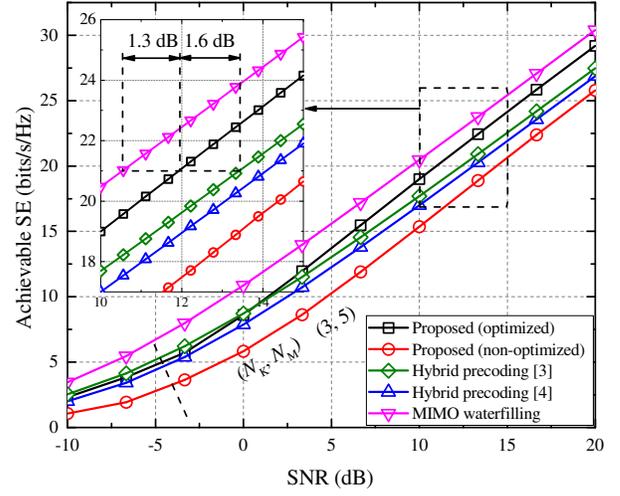


Fig. 4. Achievable SE performance yielded by various schemes with $N_T = 15$, $N_R = 10$ and $N_{\text{RF}} = 2$. The parameters (N_K, N_M) in the proposed GenSM-aided mmWave MIMO scheme is designed by solving (22).

to Fig.3, it can be seen that the proposed scheme outperforms [4] over the entire SNR range. Finally, with a target throughput of approximately 21 bits/s/Hz, it is seen that the proposed scheme outperforms [3] by approximately 1.6 dB, and is outperformed by the MIMO waterfilling capacity with a 1.3 dB performance gap.

VI. CONCLUSION

In this paper, we proposed a novel GenSM-aided mmWave MIMO scheme with hybrid precoding to improve the achievable SE of conventional mmWave MIMO schemes. A closed-form lower bound was also proposed to provide an accurate approximation to the achievable SE performance. A two-step algorithm was proposed to optimize the hybrid precoder with respect to maximizing the SE lower bound. Finally, numerical simulations are conducted to substantiate the SE gain achieved by the proposed scheme.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 61471221 and No. 61471219).

REFERENCES

- [1] R. Daniels and R. Heath, "60 GHz wireless communications: Emerging requirements and design recommendations," *IEEE Veh. Technol. Mag.*, vol. 2, no. 3, pp. 41-50, Sept. 2007.
- [2] E. Torkildson, U. Madhoo, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4150-4160, Dec. 2011.
- [3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

- [4] X. Gao, L. Dai, S. Han, C.-L. I, and R. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998-1009, Apr. 2016.
- [5] P. Liu, A. Springer, "Space shift keying for LOS communication at mmWave frequencies," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 121-124, Apr. 2015.
- [6] P. Liu, M. Di Renzo, and A. Springer, "Line-of-sight (LOS) spatial modulation (SM) for indoor mmWave communication at 60 GHz," *IEEE Trans. Wireless Commun.*, to be published.
- [7] N. Ishikawa, R. Rajashekar, S. Sugiura, and L. Hanzo, "Generalized spatial modulation based reduced-RF-chain millimeter-wave communications," *IEEE Trans. Veh. Technol.*, to be published.
- [8] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228-2241, Jul. 2008.
- [9] J. Wang, S. Jia, and J. Song, "Generalised spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1605-1615, Apr. 2012.
- [10] L. He, J. Wang, W. Ding, and J. Song, " l_∞ minimization based symbol detection for generalized space shift keying," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1109-1112, Jul. 2015.
- [11] L. He, J. Wang, and J. Song, "Spatial modulation for more spatial multiplexing: RF-chain-limited generalized spatial modulation aided mmWave MIMO with hybrid precoding," *IEEE Trans. Commun.*, submitted.
- [12] W. Zeng, C. Xiao, M. Wang, and J. Lu, "Linear precoding for finite-alphabet inputs over MIMO fading channels with statistical CSI," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3134-3148.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.